

09/19/00  
JC927 U.S. PTO

9-21-00

A/BOX/Seq 8

Express Mail No.: EK916750885US  
Docket No.: 788CIP2C

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE  
PATENT APPLICATION TRANSMITTAL UNDER 37 CFR 1.53

JC675 U.S. PTO  
09/664641  
09/19/00

BOX PATENT APPLICATION  
Assistant Commissioner for Patents  
Washington, D.C. 20231

Sir:

Transmitted herewith for filing is the patent application of

Inventor(s): Y. Tom Tang, Chenghua Liu, Ping Zhou, Vinod Asundi, Jie Zhang, Qing A. Zhao, Feiyan Ren, Aidong J. Xue, Yonghong Yang, Tom Wehrman, Radoje T. Drmanac

Title: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

1. Type of application

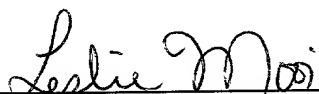
- ☒ This is a new application for a
  - ☒ Utility patent.
  - ☐ Design patent.
- ☒ This is a continuation-in-part application of prior application no. 09/577,409 filed May 18, 2000, Attorney Docket No. 788CIP, which is a continuation-in-part application of prior application no. 09/515,126 filed February 28, 2000, Attorney Docket No. 788.

2. Application Papers Enclosed

- 1 Title Page
- 112 Pages of Specification (excluding Claims, Abstract, Drawings & Sequence Listing)
- 4 Page(s) of Claims
- 1 Page(s) of Abstract
- 0 Sheet(s) of Drawings (Figs. X-X) ☐ Formal ☐ Informal
- 78 Page(s) of Sequence Listing

CERTIFICATION UNDER 37 CFR 1.10

I hereby certify that this Patent Application Transmittal and the documents referred to as enclosed therewith are being deposited with the United States Postal Service on September 19, 2000, in an envelope addressed to the Assistant Commissioner for Patents, Washington, D.C. 20231 utilizing the "Express Mail Post Office to Addressee" service of the United States Postal Service under Mailing Label No. EK916750885US

  
\_\_\_\_\_  
Leslie A. Moon

**3. Oath or Declaration**

- ☐ Enclosed
- ☐ Executed by (check all applicable boxes)
- ☐ Inventor(s)
- ☐ Legal representative of inventors(s) (37 CFR 1.42 or 1.43)
- ☐ Joint inventor or person showing a proprietary interest on behalf of inventor who refused to sign or cannot be reached
- ☐ The petition required by 37 CFR 1.47 and the statement required by 37 CFR 1.47 are enclosed. See Item 5D below for fee.
- ☒ Unexecuted – the undersigned attorney or agent is authorized to file this application on behalf of the applicant(s). An executed declaration will follow.

**4. Additional Papers Enclosed**

- ☐ Preliminary Amendment
- ☐ Information Disclosure Statement
- ☐ Declaration of Biological Deposit
- ☒ Computer readable copy of sequence listing containing nucleotide and/or amino acid sequence
- ☒ Statement Under 37 CFR § 1.821
- ☒ Paper copy of sequence listing identical to computer copy (78 pages)
- ☐ Microfiche computer program
- ☒ Verified statement claiming small entity status under 37 CFR 1.9 and 1.27
- ☐ Associate Power of Attorney
- ☐ Verified translation of a non-English patent application
- ☒ Return receipt postcard
- ☐ Other \_\_\_\_\_

**5. Priority Applications Under 35 USC 119**

Certified copies of applications from which priority under 35 USC 119 is claimed are listed below and

- ☐ are attached.
- ☐ will follow.

6. **Filing Fee Calculation (37 CFR 1.16)**

A. ☒ **Utility Application**

CLAIMS AS FILED – INCLUDING PRELIMINARY AMENDMENT (IF ANY)						
			SMALL ENTITY		OTHER THAN A SMALL ENTITY	
	NO. FILED	NO. EXTRA	RATE	FEE	RATE	FEE
BASIC FEE				\$345.00		\$690.00
TOTAL	30-20	= 10	X 9 =	\$90.00	X 18 =	\$0.00
INDEP.	3-3	= 0	X 39 =	\$0.00	X 78 =	\$0.00
<input checked="" type="checkbox"/> First Presentation of Multiple Dependent Claim			+ 130 =	\$130.00	+ 260 =	\$0.00
FILING FEE:				\$565.00	OR	\$0.00

B. ☐ **Design Application (\$155.00/\$310.00)** Filing Fee: \$ \_\_\_\_\_

C. ☐ **Plant Application (\$240.00/\$480.00)** Filing Fee: \$ \_\_\_\_\_

D. **Other fees**

☐ Recording Assignment [Fee -- \$40.00 per assignment] \$ \_\_\_\_\_

☐ Other \$ \_\_\_\_\_

**TOTAL FEES \$ 565.00**

7. **Method of Payments of Fees**

- ☐ Enclosed check
- ☒ Charge Deposit Account No. 501169. A duplicate copy of this transmittal is enclosed
- ☐ Not enclosed

8. **Deposit Account and Refund Authorization**

The Commissioner is hereby authorized to charge payment of any additional fees due or credit any overpayment to Deposit Account No. 501169. A duplicate copy of this transmittal is enclosed.

Please refund any overpayment to Hyseq, Inc. at the address below.

Please direct all future correspondence to Leslie A. Mooi at the address below.

Respectfully submitted,

Date: September 19, 2000

By:

  
\_\_\_\_\_  
Leslie A. Mooi  
Attorney for Applicants  
Registration No.: 37,047  
HYSEQ, INC.  
670 Almanor Avenue  
Sunnyvale, CA 94085  
(408) 524-8100  
(408) 524-8145 (Telefacsimile)



**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

Applicant(s) or Patentee(s): Y. Tom Tang, Chenghua Liu, Ping Zhou, Vinod Asundi, Jie Zhang, Qing A. Zhao, Feiyan Ren, Aidong J. Xue, Yonghong Yang, Tom Wehrman, Radoje T. Drmanac

Application No. or Patent No.: Not Yet Assigned

Filed or Issued: Herewith

For: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

**VERIFIED STATEMENT (DECLARATION) CLAIMING SMALL ENTITY STATUS  
(37 CFR § 1.9(f) AND 1.27(c)) - SMALL BUSINESS CONCERN**

I hereby declare that I am

- ☐ The owner of the small business concern identified below:  
☒ An official of the small business concern empowered to act on behalf of the concern identified below:

NAME OF CONCERN: HYSEQ, INC.  
ADDRESS: 670 Almanor Avenue  
Sunnyvale, CA 94085

I hereby declare that the above-identified small business concern qualifies as a small business concern as defined in 13 CFR § 121.12, and reproduced in 37 CFR § 1.9(d), for purposes of paying reduced fees under § 41(a) and (b) of Title 35, United States Code, in that the number of employees of the concern, including those of its affiliates, does not exceed 500 persons. For purposes of this statement, (1) the number of employees of the business concern is the average over the previous fiscal year of the concern of the persons employed on a full-time, part-time or temporary basis during each of the pay periods of the fiscal year, and (2) concerns are affiliates of each other when either, directly or indirectly, one concern controls or has the power to control the other, or a third party or parties controls or has the power to control both.

I hereby declare that rights under contract or law have been conveyed to, and remain with, the small business concern identified above with regard to the invention, entitled NOVEL NUCLEIC ACIDS AND POLYPEPTIDES by inventors, Y. Tom Tang, Chenghua Liu, Ping Zhou, Vinod Asundi, Jie Zhang, Qing A. Zhao, Feiyan Ren, Aidong J. Xue, Yonghong Yang, Tom Wehrman, Radoje T. Drmanac, described in

- ☒ The specification filed herewith.  
☐ Application Serial No. [ ], filed [Date].  
☐ Patent No. [ ], issued [Date].

If the rights held by the above identified small business concern are not exclusive, each individual, concern or organization having rights to the invention is listed below<sup>1</sup> and no rights to the invention are held by any person, other than the inventor, who could not qualify as a small business concern under 37 CFR § 1.9(c), or by any concern which would not qualify as a small business concern under 37 CFR § 1.9(d) or a nonprofit organization under 37 CFR § 1.9(e).

Full Name: \_\_\_\_\_

Address: \_\_\_\_\_

( ) Individual ( ) Small Business Concern ( ) Nonprofit Organization

I acknowledge the duty to file, in this application or patent, notification of any change in status resulting in loss of entitlement to small entity status prior to paying, or at the time of paying, the earliest of the issue fee or any maintenance fee due after the date on which status as a small entity is no longer appropriate (37 CFR § 1.28(b)).

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under § 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application, any patent issuing thereon, or any patent to which this verified statement is directed.

Name of person signing: James N. Fletcher

Title of person  
other than owner: Secretary

Address of person signing: HYSEQ, INC.  
670 Almanor Avenue  
Sunnyvale, CA 94085

Signature:  \_\_\_\_\_

Date: 09/19/00 \_\_\_\_\_

<sup>1</sup>NOTE: Separate verified statements are required from each named person, concern or organization having rights to the invention averring to their status as small entities. (37 CFR § 1.27)

Our Ref. No.: 788CIP2C

**NOVEL NUCLEIC ACIDS AND POLYPEPTIDES**

Express Mail Label No.: EK916750885US

# NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

## 1. CROSS REFERENCE TO RELATED APPLICATIONS

This application is a continuation-in-part application of U.S. Application Serial No. 09/577,409, filed May 18, 2000, Attorney Docket No. 788CIP, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/515,126, filed February 28, 2000, Attorney Docket No. 788, both of which are incorporated herein by reference in their entirety.

## 2. BACKGROUND OF THE INVENTION

### 2.1 TECHNICAL FIELD

The present invention provides novel polynucleotides and proteins encoded by such polynucleotides, along with uses for these polynucleotides and proteins, for example in therapeutic, diagnostic and research methods.

### 2.2 BACKGROUND

Technology aimed at the discovery of protein factors (including e.g., cytokines, such as lymphokines, interferons, CSFs, chemokines, and interleukins) has matured rapidly over the past decade. The now routine hybridization cloning and expression cloning techniques clone novel polynucleotides "directly" in the sense that they rely on information directly related to the discovered protein (i.e., partial DNA/amino acid sequence of the protein in the case of hybridization cloning; activity of the protein in the case of expression cloning). More recent "indirect" cloning techniques such as signal sequence cloning, which isolates DNA sequences based on the presence of a now well-recognized secretory leader sequence motif, as well as various PCR-based or low stringency hybridization-based cloning techniques, have advanced the state of the art by making available large numbers of DNA/amino acid sequences for proteins that are known to have biological activity, for example, by virtue of their secreted nature in the case of leader sequence cloning, by virtue of their cell or tissue source in the case of

PCR-based techniques, or by virtue of structural similarity to other genes of known biological activity.

Identified polynucleotide and polypeptide sequences have numerous applications in, for example, diagnostics, forensics, gene mapping; identification of mutations responsible for genetic disorders or other traits, to assess biodiversity, and to produce many other types of data and products dependent on DNA and amino acid sequences.

### 3. SUMMARY OF THE INVENTION

The compositions of the present invention include novel isolated polypeptides, novel isolated polynucleotides encoding such polypeptides, including recombinant DNA molecules, cloned genes or degenerate variants thereof, especially naturally occurring variants such as allelic variants, antisense polynucleotide molecules, and antibodies that specifically recognize one or more epitopes present on such polypeptides, as well as hybridomas producing such antibodies.

The compositions of the present invention additionally include vectors, including expression vectors, containing the polynucleotides of the invention, cells genetically engineered to contain such polynucleotides and cells genetically engineered to express such polynucleotides.

The present invention relates to a collection or library of at least one novel nucleic acid sequence assembled from expressed sequence tags (ESTs) isolated mainly by sequencing by hybridization (SBH), and in some cases, sequences obtained from one or more public databases. The invention relates also to the proteins encoded by such polynucleotides, along with therapeutic, diagnostic and research utilities for these polynucleotides and proteins. These nucleic acid sequences are designated as SEQ ID NO: 1 – 35 and are provided in the Sequence Listing. In the nucleic acids provided in the Sequence Listing, A is adenosine; C is cytosine; G is guanosine; T is thymine; and N is any of the four bases. In the amino acids provided in the Sequence Listing, \* corresponds to the stop codon.

The nucleic acid sequences of the present invention also include, nucleic acid sequences that hybridize to the complement of SEQ ID NO: 1 – 35 under stringent hybridization conditions; nucleic acid sequences which are allelic variants or species



the art and exemplified by Vollrath et al., Science 258:52-59 (1992), as expressed sequence tags for physical mapping of the human genome.

The isolated polynucleotides of the invention include, but are not limited to, a polynucleotide comprising any one of the nucleotide sequences set forth in the SEQ ID NO: 1-35; a polynucleotide comprising any of the full length protein coding sequences of the SEQ ID NO: 1-35; and a polynucleotide comprising any of the nucleotide sequences of the mature protein coding sequences of the SEQ ID NO: 1-35. The polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes under stringent hybridization conditions to (a) the complement of any one of the nucleotide sequences set forth in the SEQ ID NO: 1-35; (b) a nucleotide sequence encoding any one of the amino acid sequences set forth in the Sequence Listing; (c) a polynucleotide which is an allelic variant of any polynucleotides recited above; (d) a polynucleotide which encodes a species homolog (e.g. orthologs) of any of the proteins recited above; or (e) a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of any of the polypeptides comprising an amino acid sequence set forth in the Sequence Listing.

The isolated polypeptides of the invention include, but are not limited to, a polypeptide comprising any of the amino acid sequences set forth in the Sequence Listing; or the corresponding full length or mature protein. Polypeptides of the invention also include polypeptides with biological activity that are encoded by (a) any of the polynucleotides having a nucleotide sequence set forth in the SEQ ID NO: 1-35; or (b) polynucleotides that hybridize to the complement of the polynucleotides of (a) under stringent hybridization conditions. Biologically or immunologically active variants of any of the polypeptide sequences in the Sequence Listing, and "substantial equivalents" thereof (e.g., with at least about 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% amino acid sequence identity) that preferably retain biological activity are also contemplated. The polypeptides of the invention may be wholly or partially chemically synthesized but are preferably produced by recombinant means using the genetically engineered cells (e.g. host cells) of the invention.

The invention also provides compositions comprising a polypeptide of the invention. Polypeptide compositions of the invention may further comprise an acceptable carrier, such as a hydrophilic, e.g., pharmaceutically acceptable, carrier.

The invention also provides host cells transformed or transfected with a polynucleotide of the invention.

The invention also relates to methods for producing a polypeptide of the invention comprising growing a culture of the host cells of the invention in a suitable culture medium under conditions permitting expression of the desired polypeptide, and purifying the polypeptide from the culture or from the host cells. Preferred embodiments include those in which the protein produced by such process is a mature form of the protein.

Polynucleotides according to the invention have numerous applications in a variety of techniques known to those skilled in the art of molecular biology. These techniques include use as hybridization probes, use as oligomers, or primers, for PCR, use for chromosome and gene mapping, use in the recombinant production of protein, and use in generation of anti-sense DNA or RNA, their chemical analogs and the like. For example, when the expression of an mRNA is largely restricted to a particular cell or tissue type, polynucleotides of the invention can be used as hybridization probes to detect the presence of the particular cell or tissue mRNA in a sample using, *e.g.*, *in situ* hybridization.

In other exemplary embodiments, the polynucleotides are used in diagnostics as expressed sequence tags for identifying expressed genes or, as well known in the art and exemplified by Vollrath et al., Science 258:52-59 (1992), as expressed sequence tags for physical mapping of the human genome.

The polypeptides according to the invention can be used in a variety of conventional procedures and methods that are currently applied to other proteins. For example, a polypeptide of the invention can be used to generate an antibody that specifically binds the polypeptide. Such antibodies, particularly monoclonal antibodies, are useful for detecting or quantitating the polypeptide in tissue. The polypeptides of the invention can also be used as molecular weight markers, and as a food supplement.

Methods are also provided for preventing, treating, or ameliorating a medical condition which comprises the step of administering to a mammalian subject a therapeutically effective amount of a composition comprising a polypeptide of the present invention and a pharmaceutically acceptable carrier.



In particular, the polypeptides and polynucleotides of the invention can be utilized, for example, in methods for the prevention and/or treatment of disorders involving aberrant protein expression or biological activity.

The present invention further relates to methods for detecting the presence of the polynucleotides or polypeptides of the invention in a sample. Such methods can, for example, be utilized as part of prognostic and diagnostic evaluation of disorders as recited herein and for the identification of subjects exhibiting a predisposition to such conditions. The invention provides a method for detecting the polynucleotides of the invention in a sample, comprising contacting the sample with a compound that binds to and forms a complex with the polynucleotide of interest for a period sufficient to form the complex and under conditions sufficient to form a complex and detecting the complex such that if a complex is detected, the polynucleotide of interest is detected. The invention also provides a method for detecting the polypeptides of the invention in a sample comprising contacting the sample with a compound that binds to and forms a complex with the polypeptide under conditions and for a period sufficient to form the complex and detecting the formation of the complex such that if a complex is formed, the polypeptide is detected.

The invention also provides kits comprising polynucleotide probes and/or monoclonal antibodies, and optionally quantitative standards, for carrying out methods of the invention. Furthermore, the invention provides methods for evaluating the efficacy of drugs, and monitoring the progress of patients, involved in clinical trials for the treatment of disorders as recited above.

The invention also provides methods for the identification of compounds that modulate (i.e., increase or decrease) the expression or activity of the polynucleotides and/or polypeptides of the invention. Such methods can be utilized, for example, for the identification of compounds that can ameliorate symptoms of disorders as recited herein. Such methods can include, but are not limited to, assays for identifying compounds and other substances that interact with (e.g., bind to) the polypeptides of the invention. The invention provides a method for identifying a compound that binds to the polypeptides of the invention comprising contacting the compound with a polypeptide of the invention in a cell for a time sufficient to form a polypeptide/compound complex, wherein the

complex drives expression of a reporter gene sequence in the cell; and detecting the complex by detecting the reporter gene sequence expression such that if expression of the reporter gene is detected the compound the binds to a polypeptide of the invention is identified.

5           The methods of the invention also provides methods for treatment which involve the administration of the polynucleotides or polypeptides of the invention to individuals exhibiting symptoms or tendencies. In addition, the invention encompasses methods for treating diseases or disorders as recited herein comprising administering compounds and other substances that modulate the overall activity of the target gene products.

10       Compounds and other substances can effect such modulation either on the level of target gene/protein expression or target protein activity.

          The polypeptides of the present invention and the polynucleotides encoding them are also useful for the same functions known to one of skill in the art as the polypeptides and polynucleotides to which they have homology (set forth in Table 1); for which they  
15       have a signature region (as set forth in Table 3); or for which they have homology to a gene family (as set forth in Table 4). If no homology is set forth for a sequence, then the polypeptides and polynucleotides of the present invention are useful for a variety of applications, as described herein, including use in arrays for detection.

## 20           **4. DETAILED DESCRIPTION OF THE INVENTION**

### **4.1 DEFINITIONS**

          It must be noted that as used herein and in the appended claims, the singular forms “a”, “an” and “the” include plural references unless the context clearly dictates  
25       otherwise.

          The term "active" refers to those forms of the polypeptide which retain the biologic and/or immunologic activities of any naturally occurring polypeptide. According to the invention, the terms “biologically active” or “biological activity” refer to a protein or peptide having structural, regulatory or biochemical functions of a naturally occurring  
30       molecule. Likewise “immunologically active” or “immunological activity” refers to the

capability of the natural, recombinant or synthetic polypeptide to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The term "activated cells" as used in this application are those cells which are engaged in extracellular or intracellular membrane trafficking, including the export of secretory or enzymatic molecules as part of a normal or disease process.

The terms "complementary" or "complementarity" refer to the natural binding of polynucleotides by base pairing. For example, the sequence 5'-AGT-3' binds to the complementary sequence 3'-TCA-5'. Complementarity between two single-stranded molecules may be "partial" such that only some of the nucleic acids bind or it may be "complete" such that total complementarity exists between the single stranded molecules. The degree of complementarity between the nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands.

The term "embryonic stem cells (ES)" refers to a cell that can give rise to many differentiated cell types in an embryo or an adult, including the germ cells. The term "germ line stem cells (GSCs)" refers to stem cells derived from primordial stem cells that provide a steady and continuous source of germ cells for the production of gametes. The term "primordial germ cells (PGCs)" refers to a small population of cells set aside from other cell lineages particularly from the yolk sac, mesenteries, or gonadal ridges during embryogenesis that have the potential to differentiate into germ cells and other cells.

PGCs are the source from which GSCs and ES cells are derived. The PGCs, the GSCs and the ES cells are capable of self-renewal. Thus these cells not only populate the germ line and give rise to a plurality of terminally differentiated cells that comprise the adult specialized organs, but are able to regenerate themselves.

The term "expression modulating fragment," EMF, means a series of nucleotides which modulates the expression of an operably linked ORF or another EMF.

As used herein, a sequence is said to "modulate the expression of an operably linked sequence" when the expression of the sequence is altered by the presence of the EMF. EMFs include, but are not limited to, promoters, and promoter modulating sequences (inducible elements). One class of EMFs are nucleic acid fragments which induce the expression of an operably linked ORF in response to a specific regulatory factor or physiological event.

The terms "nucleotide sequence" or "nucleic acid" or "polynucleotide" or "oligonucleotide" are used interchangeably and refer to a heteropolymer of nucleotides or the sequence of these nucleotides. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA) or to any DNA-like or RNA-like material. It is contemplated that where the polynucleotide is RNA, the T (thymine) in the sequences provided herein is substituted with U (uracil). Generally, nucleic acid segments provided by this invention may be assembled from fragments of the genome and short oligonucleotide linkers, or from a series of oligonucleotides, or from individual nucleotides, to provide a synthetic nucleic acid which is capable of being expressed in a recombinant transcriptional unit comprising regulatory elements derived from a microbial or viral operon, or a eukaryotic gene.

The terms "oligonucleotide fragment" or a "polynucleotide fragment", "portion," or "segment" or "probe" or "primer" are used interchangeably and refer to a sequence of nucleotide residues which are at least about 5 nucleotides, more preferably at least about 7 nucleotides, more preferably at least about 9 nucleotides, more preferably at least about 11 nucleotides and most preferably at least about 17 nucleotides. The fragment is preferably less than about 500 nucleotides, preferably less than about 200 nucleotides, more preferably less than about 100 nucleotides, more preferably less than about 50 nucleotides and most preferably less than 30 nucleotides. Preferably the probe is from about 6 nucleotides to about 200 nucleotides, preferably from about 15 to about 50 nucleotides, more preferably from about 17 to 30 nucleotides and most preferably from about 20 to 25 nucleotides. Preferably the fragments can be used in polymerase chain reaction (PCR), various hybridization procedures or microarray procedures to identify or amplify identical or related parts of mRNA or DNA molecules. A fragment or segment may uniquely identify each polynucleotide sequence of the present invention. Preferably the fragment comprises a sequence substantially similar to any one of SEQ ID NOs:1-35.

Probes may, for example, be used to determine whether specific mRNA molecules are present in a cell or tissue or to isolate similar nucleic acid sequences from chromosomal DNA as described by Walsh et al. (Walsh, P.S. et al., 1992, PCR Methods Appl 1:241-250). They may be labeled by nick translation, Klenow fill-in reaction, PCR,

or other methods well known in the art. Probes of the present invention, their preparation and/or labeling are elaborated in Sambrook, J. et al., 1989, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, NY; or Ausubel, F.M. et al., 1989, Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, both of which are incorporated herein by reference in their entirety.

The nucleic acid sequences of the present invention also include the sequence information from the nucleic acid sequences of SEQ ID NOs: 1-35. The sequence information can be a segment of any one of SEQ ID NOs: 1-35 that uniquely identifies or represents the sequence information of that sequence of SEQ ID NO: 1-35. One such segment can be a twenty-mer nucleic acid sequence because the probability that a twenty-mer is fully matched in the human genome is 1 in 300. In the human genome, there are three billion base pairs in one set of chromosomes. Because  $4^{20}$  possible twenty-mers exist, there are 300 times more twenty-mers than there are base pairs in a set of human chromosome. Using the same analysis, the probability for a seventeen-mer to be fully matched in the human genome is approximately 1 in 5. When these segments are used in arrays for expression studies, fifteen-mer segments can be used. The probability that the fifteen-mer is fully matched in the expressed sequences is also approximately one in five because expressed sequences comprise less than approximately 5% of the entire genome sequence.

Similarly, when using sequence information for detecting a single mismatch, a segment can be a twenty-five mer. The probability that the twenty-five mer would appear in a human genome with a single mismatch is calculated by multiplying the probability for a full match ( $1 \div 4^{25}$ ) times the increased probability for mismatch at each nucleotide position ( $3 \times 25$ ). The probability that an eighteen mer with a single mismatch can be detected in an array for expression studies is approximately one in five. The probability that a twenty-mer with a single mismatch can be detected in a human genome is approximately one in five.

The term "open reading frame," ORF, means a series of nucleotide triplets coding for amino acids without any termination codons and is a sequence translatable into protein.

The terms "operably linked" or "operably associated" refer to functionally related nucleic acid sequences. For example, a promoter is operably associated or operably

linked with a coding sequence if the promoter controls the transcription of the coding sequence. While operably linked nucleic acid sequences can be contiguous and in the same reading frame, certain genetic elements e.g. repressor genes are not contiguously linked to the coding sequence but still control transcription/translation of the coding sequence.

The term “pluripotent” refers to the capability of a cell to differentiate into a number of differentiated cell types that are present in an adult organism. A pluripotent cell is restricted in its differentiation capability in comparison to a totipotent cell.

The terms “polypeptide” or “peptide” or “amino acid sequence” refer to an oligopeptide, peptide, polypeptide or protein sequence or fragment thereof and to naturally occurring or synthetic molecules. A polypeptide “fragment,” “portion,” or “segment” is a stretch of amino acid residues of at least about 5 amino acids, preferably at least about 7 amino acids, more preferably at least about 9 amino acids and most preferably at least about 17 or more amino acids. The peptide preferably is not greater than about 200 amino acids, more preferably less than 150 amino acids and most preferably less than 100 amino acids. Preferably the peptide is from about 5 to about 200 amino acids. To be active, any polypeptide must have sufficient length to display biological and/or immunological activity.

The term “naturally occurring polypeptide” refers to polypeptides produced by cells that have not been genetically engineered and specifically contemplates various polypeptides arising from post-translational modifications of the polypeptide including, but not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation and acylation.

The term “translated protein coding portion” means a sequence which encodes for the full length protein which may include any leader sequence or any processing sequence.

The term “mature protein coding sequence” means a sequence which encodes a peptide or protein without a signal or leader sequence. The peptide may have been produced by processing in the cell which removes any leader/signal sequence. The peptide may be produced synthetically or the protein may have been produced using a polynucleotide only encoding for the mature protein coding sequence.

The term "derivative" refers to polypeptides chemically modified by such techniques as ubiquitination, labeling (e.g., with radionuclides or various enzymes), covalent polymer attachment such as pegylation (derivatization with polyethylene glycol) and insertion or substitution by chemical synthesis of amino acids such as ornithine, which do not normally occur in human proteins.

The term "variant"(or "analog") refers to any polypeptide differing from naturally occurring polypeptides by amino acid insertions, deletions, and substitutions, created using, e.g., recombinant DNA techniques. Guidance in determining which amino acid residues may be replaced, added or deleted without abolishing activities of interest, may be found by comparing the sequence of the particular polypeptide with that of homologous peptides and minimizing the number of amino acid sequence changes made in regions of high homology (conserved regions) or by replacing amino acids with consensus sequence.

Alternatively, recombinant variants encoding these same or similar polypeptides may be synthesized or selected by making use of the "redundancy" in the genetic code. Various codon substitutions, such as the silent changes which produce various restriction sites, may be introduced to optimize cloning into a plasmid or viral vector or expression in a particular prokaryotic or eukaryotic system. Mutations in the polynucleotide sequence may be reflected in the polypeptide or domains of other peptides added to the polypeptide to modify the properties of any part of the polypeptide, to change characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover rate.

Preferably, amino acid "substitutions" are the result of replacing one amino acid with another amino acid having similar structural and/or chemical properties, i.e., conservative amino acid replacements. "Conservative" amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved. For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine; positively charged (basic) amino acids include arginine, lysine, and histidine; and negatively charged (acidic) amino

acids include aspartic acid and glutamic acid. "Insertions" or "deletions" are preferably in the range of about 1 to 20 amino acids, more preferably 1 to 10 amino acids. The variation allowed may be experimentally determined by systematically making insertions, deletions, or substitutions of amino acids in a polypeptide molecule using recombinant

5 DNA techniques and assaying the resulting recombinant variants for activity.

Alternatively, where alteration of function is desired, insertions, deletions or non-conservative alterations can be engineered to produce altered polypeptides. Such alterations can, for example, alter one or more of the biological functions or biochemical characteristics of the polypeptides of the invention. For example, such alterations may  
10 change polypeptide characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover rate. Further, such alterations can be selected so as to generate polypeptides that are better suited for expression, scale up and the like in the host cells chosen for expression. For example, cysteine residues can be deleted or substituted with another amino acid residue in order to eliminate disulfide bridges.

15 The terms "purified" or "substantially purified" as used herein denotes that the indicated nucleic acid or polypeptide is present in the substantial absence of other biological macromolecules, *e.g.*, polynucleotides, proteins, and the like. In one embodiment, the polynucleotide or polypeptide is purified such that it constitutes at least 95% by weight, more preferably at least 99% by weight, of the indicated biological  
20 macromolecules present (but water, buffers, and other small molecules, especially molecules having a molecular weight of less than 1000 daltons, can be present).

The term "isolated" as used herein refers to a nucleic acid or polypeptide separated from at least one other component (*e.g.*, nucleic acid or polypeptide) present with the nucleic acid or polypeptide in its natural source. In one embodiment, the nucleic  
25 acid or polypeptide is found in the presence of (if anything) only a solvent, buffer, ion, or other component normally present in a solution of the same. The terms "isolated" and "purified" do not encompass nucleic acids or polypeptides present in their natural source.

The term "recombinant," when used herein to refer to a polypeptide or protein, means that a polypeptide or protein is derived from recombinant (*e.g.*, microbial, insect,  
30 or mammalian) expression systems. "Microbial" refers to recombinant polypeptides or proteins made in bacterial or fungal (*e.g.*, yeast) expression systems. As a product,



"recombinant microbial" defines a polypeptide or protein essentially free of native endogenous substances and unaccompanied by associated native glycosylation.

Polypeptides or proteins expressed in most bacterial cultures, *e.g.*, *E. coli*, will be free of glycosylation modifications; polypeptides or proteins expressed in yeast will have a glycosylation pattern in general different from those expressed in mammalian cells.

The term "recombinant expression vehicle or vector" refers to a plasmid or phage or virus or vector, for expressing a polypeptide from a DNA (RNA) sequence. An expression vehicle can comprise a transcriptional unit comprising an assembly of (1) a genetic element or elements having a regulatory role in gene expression, for example, promoters or enhancers, (2) a structural or coding sequence which is transcribed into mRNA and translated into protein, and (3) appropriate transcription initiation and termination sequences. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where recombinant protein is expressed without a leader or transport sequence, it may include an amino terminal methionine residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

The term "recombinant expression system" means host cells which have stably integrated a recombinant transcriptional unit into chromosomal DNA or carry the recombinant transcriptional unit extrachromosomally. Recombinant expression systems as defined herein will express heterologous polypeptides or proteins upon induction of the regulatory elements linked to the DNA segment or synthetic gene to be expressed. This term also means host cells which have stably integrated a recombinant genetic element or elements having a regulatory role in gene expression, for example, promoters or enhancers. Recombinant expression systems as defined herein will express polypeptides or proteins endogenous to the cell upon induction of the regulatory elements linked to the endogenous DNA segment or gene to be expressed. The cells can be prokaryotic or eukaryotic.

The term "secreted" includes a protein that is transported across or through a membrane, including transport as a result of signal sequences in its amino acid sequence when it is expressed in a suitable host cell. "Secreted" proteins include without limitation

proteins secreted wholly (e.g., soluble proteins) or partially (e.g., receptors) from the cell in which they are expressed. "Secreted" proteins also include without limitation proteins that are transported across the membrane of the endoplasmic reticulum. "Secreted" proteins are also intended to include proteins containing non-typical signal sequences (e.g. Interleukin-1 Beta, see Krasney, P.A. and Young, P.R. (1992) Cytokine 4(2):134-143) and factors released from damaged cells (e.g. Interleukin-1 Receptor Antagonist, see Arend, W.P. et. al. (1998) Annu. Rev. Immunol. 16:27-55)

Where desired, an expression vector may be designed to contain a "signal or leader sequence" which will direct the polypeptide through the membrane of a cell. Such a sequence may be naturally present on the polypeptides of the present invention or provided from heterologous protein sources by recombinant DNA techniques.

The term "stringent" is used to refer to conditions that are commonly understood in the art as stringent. Stringent conditions can include highly stringent conditions (i.e., hybridization to filter-bound DNA in 0.5 M NaHPO<sub>4</sub>, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.1X SSC/0.1% SDS at 68°C), and moderately stringent conditions (i.e., washing in 0.2X SSC/0.1% SDS at 42°C). Other exemplary hybridization conditions are described herein in the examples.

In instances of hybridization of deoxyoligonucleotides, additional exemplary stringent hybridization conditions include washing in 6X SSC/0.05% sodium pyrophosphate at 37°C (for 14-base oligonucleotides), 48°C (for 17-base oligos), 55°C (for 20-base oligonucleotides), and 60°C (for 23-base oligonucleotides).

As used herein, "substantially equivalent" can refer both to nucleotide and amino acid sequences, for example a mutant sequence, that varies from a reference sequence by one or more substitutions, deletions, or additions, the net effect of which does not result in an adverse functional dissimilarity between the reference and subject sequences. Typically, such a substantially equivalent sequence varies from one of those listed herein by no more than about 35% (i.e., the number of individual residue substitutions, additions, and/or deletions in a substantially equivalent sequence, as compared to the corresponding reference sequence, divided by the total number of residues in the substantially equivalent sequence is about 0.35 or less). Such a sequence is said to have 65% sequence identity to the listed sequence. In one embodiment, a substantially

equivalent, *e.g.*, mutant, sequence of the invention varies from a listed sequence by no more than 30% (70% sequence identity); in a variation of this embodiment, by no more than 25% (75% sequence identity); and in a further variation of this embodiment, by no more than 20% (80% sequence identity) and in a further variation of this embodiment, by no more than 10% (90% sequence identity) and in a further variation of this embodiment, by no more than 5% (95% sequence identity). Substantially equivalent, *e.g.*, mutant, amino acid sequences according to the invention preferably have at least 80% sequence identity with a listed amino acid sequence, more preferably at least 90% sequence identity. Substantially equivalent nucleotide sequences of the invention can have lower percent sequence identities, taking into account, for example, the redundancy or degeneracy of the genetic code. Preferably, nucleotide sequence has at least about 65% identity, more preferably at least about 75% identity, and most preferably at least about 95% identity. For the purposes of the present invention, sequences having substantially equivalent biological activity and substantially equivalent expression characteristics are considered substantially equivalent. For the purposes of determining equivalence, truncation of the mature sequence (*e.g.*, via a mutation which creates a spurious stop codon) should be disregarded. Sequence identity may be determined, *e.g.*, using the Jotun Hein method (Hein, J. (1990) *Methods Enzymol.* 183:626-645). Identity between sequences can also be determined by other methods known in the art, *e.g.* by varying hybridization conditions.

The term "totipotent" refers to the capability of a cell to differentiate into all of the cell types of an adult organism.

The term "transformation" means introducing DNA into a suitable host cell so that the DNA is replicable, either as an extrachromosomal element, or by chromosomal integration. The term "transfection" refers to the taking up of an expression vector by a suitable host cell, whether or not any coding sequences are in fact expressed. The term "infection" refers to the introduction of nucleic acids into a suitable host cell by use of a virus or viral vector.

As used herein, an "uptake modulating fragment," UMF, means a series of nucleotides which mediate the uptake of a linked DNA fragment into a cell. UMFs can be readily identified using known UMFs as a target sequence or target motif with the

computer-based systems described below. The presence and activity of a UMF can be confirmed by attaching the suspected UMF to a marker sequence. The resulting nucleic acid molecule is then incubated with an appropriate host under appropriate conditions and the uptake of the marker sequence is determined. As described above, a UMF will increase the frequency of uptake of a linked marker sequence.

Each of the above terms is meant to encompass all that is described for each, unless the context dictates otherwise.

## 4.2 NUCLEIC ACIDS OF THE INVENTION

Nucleotide sequences of the invention are set forth in the Sequence Listing.

The isolated polynucleotides of the invention include a polynucleotide comprising the nucleotide sequences of the SEQ ID NO: 1 – 35; a polynucleotide encoding any one of the peptide sequences of SEQ ID NO:1 – 35; and a polynucleotide comprising the nucleotide sequence encoding the mature protein coding sequence of the polynucleotides of any one of SEQ ID NO: 1 - 35. The polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes under stringent conditions to (a) the complement of any of the nucleotides sequences of the SEQ ID NO: 1 – 35; (b) nucleotide sequences encoding any one of the amino acid sequences set forth in the Sequence Listing; (c) a polynucleotide which is an allelic variant of any polynucleotide recited above; (d) a polynucleotide which encodes a species homolog of any of the proteins recited above; or (e) a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of the polypeptides of SEQ ID NO: 1- 35. Domains of interest may depend on the nature of the encoded polypeptide; e.g., domains in receptor-like polypeptides include ligand-binding, extracellular, transmembrane, or cytoplasmic domains, or combinations thereof; domains in immunoglobulin-like proteins include the variable immunoglobulin-like domains; domains in enzyme-like polypeptides include catalytic and substrate binding domains; and domains in ligand polypeptides include receptor-binding domains.

The polynucleotides of the invention include naturally occurring or wholly or partially synthetic DNA, e.g., cDNA and genomic DNA, and RNA, e.g., mRNA. The

polynucleotides may include all of the coding region of the cDNA or may represent a portion of the coding region of the cDNA.

The present invention also provides genes corresponding to the cDNA sequences disclosed herein. The corresponding genes can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include the preparation of probes or primers from the disclosed sequence information for identification and/or amplification of genes in appropriate genomic libraries or other sources of genomic materials. Further 5' and 3' sequence can be obtained using methods known in the art. For example, full length cDNA or genomic DNA that corresponds to any of the polynucleotides of the SEQ ID NO: 1 – 35 can be obtained by screening appropriate cDNA or genomic DNA libraries under suitable hybridization conditions using any of the polynucleotides of the SEQ ID NO: 1 - 35 or a portion thereof as a probe. Alternatively, the polynucleotides of the SEQ ID NO: 1 - 35 may be used as the basis for suitable primer(s) that allow identification and/or amplification of genes in appropriate genomic DNA or cDNA libraries.

The nucleic acid sequences of the invention can be assembled from ESTs and sequences (including cDNA and genomic sequences) obtained from one or more public databases, such as dbEST, gbpri, and UniGene. The EST sequences can provide identifying sequence information, representative fragment or segment information, or novel segment information for the full-length gene.

The polynucleotides of the invention also provide polynucleotides including nucleotide sequences that are substantially equivalent to the polynucleotides recited above. Polynucleotides according to the invention can have, *e.g.*, at least about 65%, at least about 70%, at least about 75%, at least about 80%, more typically at least about 90%, and even more typically at least about 95%, sequence identity to a polynucleotide recited above.

Included within the scope of the nucleic acid sequences of the invention are nucleic acid sequence fragments that hybridize under stringent conditions to any of the nucleotide sequences of the SEQ ID NO: 1 - 35, or complements thereof, which fragment is greater than about 5 nucleotides, preferably 7 nucleotides, more preferably greater than 9 nucleotides and most preferably greater than 17 nucleotides. Fragments of, *e.g.* 15, 17, or 20 nucleotides or more that are selective for (*i.e.* specifically hybridize to any one of

the polynucleotides of the invention) are contemplated. Probes capable of specifically hybridizing to a polynucleotide can differentiate polynucleotide sequences of the invention from other polynucleotide sequences in the same family of genes or can differentiate human genes from genes of other species, and are preferably based on unique nucleotide sequences.

The sequences falling within the scope of the present invention are not limited to these specific sequences, but also include allelic and species variations thereof. Allelic and species variations can be routinely determined by comparing the sequence provided in SEQ ID NO: 1 - 35, a representative fragment thereof, or a nucleotide sequence at least 90% identical, preferably 95% identical, to SEQ ID NOs: 1 - 35 with a sequence from another isolate of the same species. Furthermore, to accommodate codon variability, the invention includes nucleic acid molecules coding for the same amino acid sequences as do the specific ORFs disclosed herein. In other words, in the coding region of an ORF, substitution of one codon for another codon that encodes the same amino acid is expressly contemplated.

The nearest neighbor or homology result for the nucleic acids of the present invention, including SEQ ID NOs: 1 - 35, can be obtained by searching a database using an algorithm or a program. Preferably, a BLAST which stands for Basic Local Alignment Search Tool is used to search for local sequence alignments (Altshul, S.F. J Mol. Evol. 36 290-300 (1993) and Altschul S.F. et al. J. Mol. Biol. 21:403-410 (1990)). Alternatively a FASTA version 3 search against Genpept, using Fastxy algorithm.

Species homologs (or orthologs) of the disclosed polynucleotides and proteins are also provided by the present invention. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source from the desired species.

The invention also encompasses allelic variants of the disclosed polynucleotides or proteins; that is, naturally-occurring alternative forms of the isolated polynucleotide which also encode proteins which are identical, homologous or related to that encoded by the polynucleotides.

The nucleic acid sequences of the invention are further directed to sequences which encode variants of the described nucleic acids. These amino acid sequence variants may be prepared by methods known in the art by introducing appropriate

nucleotide changes into a native or variant polynucleotide. There are two variables in the construction of amino acid sequence variants: the location of the mutation and the nature of the mutation. Nucleic acids encoding the amino acid sequence variants are preferably constructed by mutating the polynucleotide to encode an amino acid sequence that does not occur in nature. These nucleic acid alterations can be made at sites that differ in the nucleic acids from different species (variable positions) or in highly conserved regions (constant regions). Sites at such locations will typically be modified in series, *e.g.*, by substituting first with conservative choices (*e.g.*, hydrophobic amino acid to a different hydrophobic amino acid) and then with more distant choices (*e.g.*, hydrophobic amino acid to a charged amino acid), and then deletions or insertions may be made at the target site. Amino acid sequence deletions generally range from about 1 to 30 residues, preferably about 1 to 10 residues, and are typically contiguous. Amino acid insertions include amino- and/or carboxyl-terminal fusions ranging in length from one to one hundred or more residues, as well as intrasequence insertions of single or multiple amino acid residues. Intrasequence insertions may range generally from about 1 to 10 amino residues, preferably from 1 to 5 residues. Examples of terminal insertions include the heterologous signal sequences necessary for secretion or for intracellular targeting in different host cells and sequences such as FLAG or poly-histidine sequences useful for purifying the expressed protein.

In a preferred method, polynucleotides encoding the novel amino acid sequences are changed via site-directed mutagenesis. This method uses oligonucleotide sequences to alter a polynucleotide to encode the desired amino acid variant, as well as sufficient adjacent nucleotides on both sides of the changed amino acid to form a stable duplex on either side of the site of being changed. In general, the techniques of site-directed mutagenesis are well known to those of skill in the art and this technique is exemplified by publications such as, Edelman et al., *DNA* 2:183 (1983). A versatile and efficient method for producing site-specific changes in a polynucleotide sequence was published by Zoller and Smith, *Nucleic Acids Res.* 10:6487-6500 (1982). PCR may also be used to create amino acid sequence variants of the novel nucleic acids. When small amounts of template DNA are used as starting material, primer(s) that differs slightly in sequence from the corresponding region in the template DNA can generate the desired amino acid

variant. PCR amplification results in a population of product DNA fragments that differ from the polynucleotide template encoding the polypeptide at the position specified by the primer. The product DNA fragments replace the corresponding region in the plasmid and this gives a polynucleotide encoding the desired amino acid variant.

5 A further technique for generating amino acid variants is the cassette mutagenesis technique described in Wells et al., *Gene* 34:315 (1985); and other mutagenesis techniques well known in the art, such as, for example, the techniques in Sambrook et al., supra, and *Current Protocols in Molecular Biology*, Ausubel et al. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the  
10 same or a functionally equivalent amino acid sequence may be used in the practice of the invention for the cloning and expression of these novel nucleic acids. Such DNA sequences include those which are capable of hybridizing to the appropriate novel nucleic acid sequence under stringent conditions.

Polynucleotides encoding preferred polypeptide truncations of the invention can  
15 be used to generate polynucleotides encoding chimeric or fusion proteins comprising one or more domains of the invention and heterologous protein sequences.

The polynucleotides of the invention additionally include the complement of any of the polynucleotides recited above. The polynucleotide can be DNA (genomic, cDNA, amplified, or synthetic) or RNA. Methods and algorithms for obtaining such  
20 polynucleotides are well known to those of skill in the art and can include, for example, methods for determining hybridization conditions that can routinely isolate polynucleotides of the desired sequence identities.

In accordance with the invention, polynucleotide sequences comprising the mature protein coding sequences corresponding to any one of SEQ ID NO: 1-35, or  
25 functional equivalents thereof, may be used to generate recombinant DNA molecules that direct the expression of that nucleic acid, or a functional equivalent thereof, in appropriate host cells. Also included are the cDNA inserts of any of the clones identified herein.

A polynucleotide according to the invention can be joined to any of a variety of  
30 other nucleotide sequences by well-established recombinant DNA techniques (see Sambrook J et al. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor



Laboratory, NY). Useful nucleotide sequences for joining to polynucleotides include an assortment of vectors, e.g., plasmids, cosmids, lambda phage derivatives, phagemids, and the like, that are well known in the art. Accordingly, the invention also provides a vector including a polynucleotide of the invention and a host cell containing the polynucleotide.

5 In general, the vector contains an origin of replication functional in at least one organism, convenient restriction endonuclease sites, and a selectable marker for the host cell.

Vectors according to the invention include expression vectors, replication vectors, probe generation vectors, and sequencing vectors. A host cell according to the invention can be a prokaryotic or eukaryotic cell and can be a unicellular organism or part of a

10 multicellular organism.

The present invention further provides recombinant constructs comprising a nucleic acid having any of the nucleotide sequences of the SEQ ID NOs: 1 - 35 or a fragment thereof or any other polynucleotides of the invention. In one embodiment, the recombinant constructs of the present invention comprise a vector, such as a plasmid or  
15 viral vector, into which a nucleic acid having any of the nucleotide sequences of the SEQ ID NOs: 1 - 35 or a fragment thereof is inserted, in a forward or reverse orientation. In the case of a vector comprising one of the ORFs of the present invention, the vector may further comprise regulatory sequences, including for example, a promoter, operably linked to the ORF. Large numbers of suitable vectors and promoters are known to those  
20 of skill in the art and are commercially available for generating the recombinant constructs of the present invention. The following vectors are provided by way of example. Bacterial: pBs, phagescript, PsiX174, pBluescript SK, pBs KS, pNH8a, pNH16a, pNH18a, pNH46a (Stratagene); pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia). Eukaryotic: pWLneo, pSV2cat, pOG44, PXTI, pSG (Stratagene)  
25 pSVK3, pBPV, pMSG, pSVL (Pharmacia).

The isolated polynucleotide of the invention may be operably linked to an expression control sequence such as the pMT2 or pED expression vectors disclosed in Kaufman et al., *Nucleic Acids Res.* 19, 4485-4490 (1991), in order to produce the protein recombinantly. Many suitable expression control sequences are known in the art.

30 General methods of expressing recombinant proteins are also known and are exemplified in R. Kaufman, *Methods in Enzymology* 185, 537-566 (1990). As defined herein

"operably linked" means that the isolated polynucleotide of the invention and an expression control sequence are situated within a vector or cell in such a way that the protein is expressed by a host cell which has been transformed (transfected) with the ligated polynucleotide/expression control sequence.

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda PR, and trc. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art. Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, *e.g.*, the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), a-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an amino terminal identification peptide imparting desired characteristics, *e.g.*, stabilization or simplified purification of expressed recombinant product. Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include *E. coli*, *Bacillus subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*, although others may also be employed as a matter of choice.

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM 1 (Promega Biotech, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed. Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced or derepressed by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period. Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Polynucleotides of the invention can also be used to induce immune responses. For example, as described in Fan et al., *Nat. Biotech.* 17:870-872 (1999), incorporated herein by reference, nucleic acid sequences encoding a polypeptide may be used to generate antibodies against the encoded polypeptide following topical administration of naked plasmid DNA or following injection, and preferably intramuscular injection of the DNA. The nucleic acid sequences are preferably inserted in a recombinant expression vector and may be in the form of naked DNA.

### 4.3 HOSTS

The present invention further provides host cells genetically engineered to contain the polynucleotides of the invention. For example, such host cells may contain nucleic acids of the invention introduced into the host cell using known transformation, transfection or infection methods. The present invention still further provides host cells genetically engineered to express the polynucleotides of the invention, wherein such polynucleotides are in operative association with a regulatory sequence heterologous to the host cell which drives expression of the polynucleotides in the cell.

Knowledge of nucleic acid sequences allows for modification of cells to permit, or increase, expression of endogenous polypeptide. Cells can be modified (e.g., by homologous recombination) to provide increased polypeptide expression by replacing, in

whole or in part, the naturally occurring promoter with all or part of a heterologous promoter so that the cells express the polypeptide at higher levels. The heterologous promoter is inserted in such a manner that it is operatively linked to the encoding sequences. See, for example, PCT International Publication No. WO94/12650, PCT International Publication No. WO92/20808, and PCT International Publication No. WO91/09955. It is also contemplated that, in addition to heterologous promoter DNA, amplifiable marker DNA (e.g., *ada*, *dhfr*, and the multifunctional CAD gene which encodes carbamyl phosphate synthase, aspartate transcarbamylase, and dihydroorotase) and/or intron DNA may be inserted along with the heterologous promoter DNA. If linked to the coding sequence, amplification of the marker DNA by standard selection methods results in co-amplification of the desired protein coding sequences in the cells.

The host cell can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the recombinant construct into the host cell can be effected by calcium phosphate transfection, DEAE, dextran mediated transfection, or electroporation (Davis, L. et al., *Basic Methods in Molecular Biology* (1986)). The host cells containing one of the polynucleotides of the invention, can be used in conventional manners to produce the gene product encoded by the isolated fragment (in the case of an ORF) or can be used to produce a heterologous protein under the control of the EMF.

Any host/vector system can be used to express one or more of the ORFs of the present invention. These include, but are not limited to, eukaryotic hosts such as HeLa cells, Cv-1 cell, COS cells, 293 cells, and Sf9 cells, as well as prokaryotic host such as *E. coli* and *B. subtilis*. The most preferred cells are those which do not normally express the particular polypeptide or protein or which expresses the polypeptide or protein at low natural level. Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, et al., in *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor, New York (1989), the disclosure of which is hereby incorporated by reference.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, Cell 23:175 (1981). Other cell lines capable of expressing a compatible vector are, for example, the C127, monkey

5 COS cells, Chinese Hamster Ovary (CHO) cells, human kidney 293 cells, human epidermal A431 cells, human Colo205 cells, 3T3 cells, CV-1 cells, other transformed primate cell lines, normal diploid cells, cell strains derived from *in vitro* culture of primary tissue, primary explants, HeLa cells, mouse L cells, BHK, HL-60, U937, HaK or Jurkat cells. Mammalian expression vectors will comprise an origin of replication, a

10 suitable promoter and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements. Recombinant polypeptides

15 and proteins produced in bacterial culture are usually isolated by initial extraction from cell pellets, followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps. Microbial cells employed in

20 expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents.

Alternatively, it may be possible to produce the protein in lower eukaryotes such as yeast or insects or in prokaryotes such as bacteria. Potentially suitable yeast strains include *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Kluyveromyces* strains,

25 *Candida*, or any yeast strain capable of expressing heterologous proteins. Potentially suitable bacterial strains include *Escherichia coli*, *Bacillus subtilis*, *Salmonella typhimurium*, or any bacterial strain capable of expressing heterologous proteins. If the protein is made in yeast or bacteria, it may be necessary to modify the protein produced therein, for example by phosphorylation or glycosylation of the appropriate sites, in order

30 to obtain the functional protein. Such covalent attachments may be accomplished using known chemical or enzymatic methods.

In another embodiment of the present invention, cells and tissues may be engineered to express an endogenous gene comprising the polynucleotides of the invention under the control of inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene targeting can be used to replace a gene's existing regulatory region with a regulatory sequence isolated from a different gene or a novel regulatory sequence synthesized by genetic engineering methods. Such regulatory sequences may be comprised of promoters, enhancers, scaffold-attachment regions, negative regulatory elements, transcriptional initiation sites, regulatory protein binding sites or combinations of said sequences. Alternatively, sequences which affect the structure or stability of the RNA or protein produced may be replaced, removed, added, or otherwise modified by targeting. These sequence include polyadenylation signals, mRNA stability elements, splice sites, leader sequences for enhancing or modifying transport or secretion properties of the protein, or other sequences which alter or improve the function or stability of protein or RNA molecules.

The targeting event may be a simple insertion of the regulatory sequence, placing the gene under the control of the new regulatory sequence, *e.g.*, inserting a new promoter or enhancer or both upstream of a gene. Alternatively, the targeting event may be a simple deletion of a regulatory element, such as the deletion of a tissue-specific negative regulatory element. Alternatively, the targeting event may replace an existing element; for example, a tissue-specific enhancer can be replaced by an enhancer that has broader or different cell-type specificity than the naturally occurring elements. Here, the naturally occurring sequences are deleted and new sequences are added. In all cases, the identification of the targeting event may be facilitated by the use of one or more selectable marker genes that are contiguous with the targeting DNA, allowing for the selection of cells in which the exogenous DNA has integrated into the host cell genome. The identification of the targeting event may also be facilitated by the use of one or more marker genes exhibiting the property of negative selection, such that the negatively selectable marker is linked to the exogenous DNA, but configured such that the negatively selectable marker flanks the targeting sequence, and such that a correct homologous recombination event with sequences in the host cell genome does not result

in the stable integration of the negatively selectable marker. Markers useful for this purpose include the Herpes Simplex Virus thymidine kinase (TK) gene or the bacterial xanthine-guanine phosphoribosyl-transferase (gpt) gene.

The gene targeting or gene activation techniques which can be used in accordance with this aspect of the invention are more particularly described in U.S. Patent No. 5,272,071 to Chappel; U.S. Patent No. 5,578,461 to Sherwin et al.; International Application No. PCT/US92/09627 (WO93/09222) by Selden et al.; and International Application No. PCT/US90/06436 (WO91/06667) by Skoultchi et al., each of which is incorporated by reference herein in its entirety.

#### 4.4 POLYPEPTIDES OF THE INVENTION

The isolated polypeptides of the invention include, but are not limited to, a polypeptide comprising: the amino acid sequences set forth as any one of SEQ ID NO: 1-35 or an amino acid sequence encoded by any one of the nucleotide sequences SEQ ID NOs: 1 - 35 or the corresponding full length or mature protein. Polypeptides of the invention also include polypeptides preferably with biological or immunological activity that are encoded by: (a) a polynucleotide having any one of the nucleotide sequences set forth in the SEQ ID NOs: 1 – 35 or (b) polynucleotides encoding any one of the amino acid sequences set forth as SEQ ID NO: 1-35 or (c) polynucleotides that hybridize to the complement of the polynucleotides of either (a) or (b) under stringent hybridization conditions. The invention also provides biologically active or immunologically active variants of any of the amino acid sequences set forth as SEQ ID NO: 1-35 or the corresponding full length or mature protein; and “substantial equivalents” thereof (e.g., with at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, typically at least about 95%, more typically at least about 98%, or most typically at least about 99% amino acid identity) that retain biological activity. Polypeptides encoded by allelic variants may have a similar, increased, or decreased activity compared to polypeptides comprising SEQ ID NO: 1-35.

Fragments of the proteins of the present invention which are capable of exhibiting biological activity are also encompassed by the present invention. Fragments of the protein may be in linear form or they may be cyclized using known methods, for

example, as described in H. U. Saragovi, et al., Bio/Technology 10, 773-778 (1992) and in R. S. McDowell, et al., J. Amer. Chem. Soc. 114, 9245-9253 (1992), both of which are incorporated herein by reference. Such fragments may be fused to carrier molecules such as immunoglobulins for many purposes, including increasing the valency of protein binding sites.

The present invention also provides both full-length and mature forms (for example, without a signal sequence or precursor sequence) of the disclosed proteins. The protein coding sequence is identified in the sequence listing by translation of the disclosed nucleotide sequences. The mature form of such protein may be obtained by expression of a full-length polynucleotide in a suitable mammalian cell or other host cell. The sequence of the mature form of the protein is also determinable from the amino acid sequence of the full-length form. Where proteins of the present invention are membrane bound, soluble forms of the proteins are also provided. In such forms, part or all of the regions causing the proteins to be membrane bound are deleted so that the proteins are fully secreted from the cell in which it is expressed.

Protein compositions of the present invention may further comprise an acceptable carrier, such as a hydrophilic, *e.g.*, pharmaceutically acceptable, carrier.

The present invention further provides isolated polypeptides encoded by the nucleic acid fragments of the present invention or by degenerate variants of the nucleic acid fragments of the present invention. By "degenerate variant" is intended nucleotide fragments which differ from a nucleic acid fragment of the present invention (*e.g.*, an ORF) by nucleotide sequence but, due to the degeneracy of the genetic code, encode an identical polypeptide sequence. Preferred nucleic acid fragments of the present invention are the ORFs that encode proteins.

A variety of methodologies known in the art can be utilized to obtain any one of the isolated polypeptides or proteins of the present invention. At the simplest level, the amino acid sequence can be synthesized using commercially available peptide synthesizers. The synthetically-constructed protein sequences, by virtue of sharing primary, secondary or tertiary structural and/or conformational characteristics with proteins may possess biological properties in common therewith, including protein activity. This technique is particularly useful in producing small peptides and fragments



of larger polypeptides. Fragments are useful, for example, in generating antibodies against the native polypeptide. Thus, they may be employed as biologically active or immunological substitutes for natural, purified proteins in screening of therapeutic compounds and in immunological processes for the development of antibodies.

5           The polypeptides and proteins of the present invention can alternatively be purified from cells which have been altered to express the desired polypeptide or protein. As used herein, a cell is said to be altered to express a desired polypeptide or protein when the cell, through genetic manipulation, is made to produce a polypeptide or protein which it normally does not produce or which the cell normally produces at a lower level.

10          One skilled in the art can readily adapt procedures for introducing and expressing either recombinant or synthetic sequences into eukaryotic or prokaryotic cells in order to generate a cell which produces one of the polypeptides or proteins of the present invention.

15           The invention also relates to methods for producing a polypeptide comprising growing a culture of host cells of the invention in a suitable culture medium, and purifying the protein from the cells or the culture in which the cells are grown. For example, the methods of the invention include a process for producing a polypeptide in which a host cell containing a suitable expression vector that includes a polynucleotide of the invention is cultured under conditions that allow expression of the encoded  
20          polypeptide. The polypeptide can be recovered from the culture, conveniently from the culture medium, or from a lysate prepared from the host cells and further purified. Preferred embodiments include those in which the protein produced by such process is a full length or mature form of the protein.

25           In an alternative method, the polypeptide or protein is purified from bacterial cells which naturally produce the polypeptide or protein. One skilled in the art can readily follow known methods for isolating polypeptides and proteins in order to obtain one of the isolated polypeptides or proteins of the present invention. These include, but are not limited to, immunochromatography, HPLC, size-exclusion chromatography, ion-exchange chromatography, and immuno-affinity chromatography. See, *e.g.*, Scopes,  
30          *Protein Purification: Principles and Practice*, Springer-Verlag (1994); Sambrook, et al., in *Molecular Cloning: A Laboratory Manual*; Ausubel et al., *Current Protocols in*

*Molecular Biology*. Polypeptide fragments that retain biological/immunological activity include fragments comprising greater than about 100 amino acids, or greater than about 200 amino acids, and fragments that encode specific protein domains.

The purified polypeptides can be used in *in vitro* binding assays which are well known in the art to identify molecules which bind to the polypeptides. These molecules include but are not limited to, for e.g., small molecules, molecules from combinatorial libraries, antibodies or other proteins. The molecules identified in the binding assay are then tested for antagonist or agonist activity in *in vivo* tissue culture or animal models that are well known in the art. In brief, the molecules are titrated into a plurality of cell cultures or animals and then tested for either cell/animal death or prolonged survival of the animal/cells.

In addition, the peptides of the invention or molecules capable of binding to the peptides may be complexed with toxins, e.g., ricin or cholera, or with other compounds that are toxic to cells. The toxin-binding molecule complex is then targeted to a tumor or other cell by the specificity of the binding molecule for SEQ ID NO: 1-35.

The protein of the invention may also be expressed as a product of transgenic animals, e.g., as a component of the milk of transgenic cows, goats, pigs, or sheep which are characterized by somatic or germ cells containing a nucleotide sequence encoding the protein.

The proteins provided herein also include proteins characterized by amino acid sequences similar to those of purified proteins but into which modification are naturally provided or deliberately engineered. For example, modifications, in the peptide or DNA sequence, can be made by those skilled in the art using known techniques. Modifications of interest in the protein sequences may include the alteration, substitution, replacement, insertion or deletion of a selected amino acid residue in the coding sequence. For example, one or more of the cysteine residues may be deleted or replaced with another amino acid to alter the conformation of the molecule. Techniques for such alteration, substitution, replacement, insertion or deletion are well known to those skilled in the art (see, e.g., U.S. Pat. No. 4,518,584). Preferably, such alteration, substitution, replacement, insertion or deletion retains the desired activity of the protein. Regions of the protein that are important for the protein function can be determined by various methods known in

the art including the alanine-scanning method which involved systematic substitution of single or strings of amino acids with alanine, followed by testing the resulting alanine-containing variant for biological activity. This type of analysis determines the importance of the substituted amino acid(s) in biological activity. Regions of the protein that are important for protein function may be determined by the eMATRIX program.

Other fragments and derivatives of the sequences of proteins which would be expected to retain protein activity in whole or in part and are useful for screening or other immunological methodologies may also be easily made by those skilled in the art given the disclosures herein. Such modifications are encompassed by the present invention.

The protein may also be produced by operably linking the isolated polynucleotide of the invention to suitable control sequences in one or more insect expression vectors, and employing an insect expression system. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, e.g., Invitrogen, San Diego, Calif., U.S.A. (the MaxBat™ kit), and such methods are well known in the art, as described in Summers and Smith, Texas Agricultural Experiment Station Bulletin No. 1555 (1987), incorporated herein by reference. As used herein, an insect cell capable of expressing a polynucleotide of the present invention is "transformed."

The protein of the invention may be prepared by culturing transformed host cells under culture conditions suitable to express the recombinant protein. The resulting expressed protein may then be purified from such culture (*i.e.*, from culture medium or cell extracts) using known purification processes, such as gel filtration and ion exchange chromatography. The purification of the protein may also include an affinity column containing agents which will bind to the protein; one or more column steps over such affinity resins as concanavalin A-agarose, heparin-toyopearl™ or Cibacrom blue 3GA Sepharose™; one or more steps involving hydrophobic interaction chromatography using such resins as phenyl ether, butyl ether, or propyl ether; or immunoaffinity chromatography.

Alternatively, the protein of the invention may also be expressed in a form which will facilitate purification. For example, it may be expressed as a fusion protein, such as those of maltose binding protein (MBP), glutathione-S-transferase (GST) or thioredoxin

(TRX), or as a His tag. Kits for expression and purification of such fusion proteins are commercially available from New England BioLab (Beverly, Mass.), Pharmacia (Piscataway, N.J.) and Invitrogen, respectively. The protein can also be tagged with an epitope and subsequently purified by using a specific antibody directed to such epitope.

- 5 One such epitope ("FLAG®") is commercially available from Kodak (New Haven, Conn.).

Finally, one or more reverse-phase high performance liquid chromatography (RP-HPLC) steps employing hydrophobic RP-HPLC media, *e.g.*, silica gel having pendant methyl or other aliphatic groups, can be employed to further purify the protein. Some or  
10 all of the foregoing purification steps, in various combinations, can also be employed to provide a substantially homogeneous isolated recombinant protein. The protein thus purified is substantially free of other mammalian proteins and is defined in accordance with the present invention as an "isolated protein."

The polypeptides of the invention include analogs (variants). This embraces  
15 fragments, as well as peptides in which one or more amino acids has been deleted, inserted, or substituted. Also, analogs of the polypeptides of the invention embrace fusions of the polypeptides or modifications of the polypeptides of the invention, wherein the polypeptide or analog is fused to another moiety or moieties, *e.g.*, targeting moiety or another therapeutic agent. Such analogs may exhibit improved properties such as activity  
20 and/or stability. Examples of moieties which may be fused to the polypeptide or an analog include, for example, targeting moieties which provide for the delivery of polypeptide to pancreatic cells, *e.g.*, antibodies to pancreatic cells, antibodies to immune cells such as T-cells, monocytes, dendritic cells, granulocytes, etc., as well as receptor and ligands expressed on pancreatic or immune cells. Other moieties which may be  
25 fused to the polypeptide include therapeutic agents which are used for treatment, for example, immunosuppressive drugs such as cyclosporin, SK506, azathioprine, CD3 antibodies and steroids. Also, polypeptides may be fused to immune modulators, and other cytokines such as alpha or beta interferon.

#### 30 4.4.1 DETERMINING POLYPEPTIDE AND POLYNUCLEOTIDE IDENTITY AND SIMILARITY

Preferred identity and/or similarity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in computer programs including, but are not limited to, the GCG program package, including GAP (Devereux, J., et al., Nucleic Acids Research 12(1):387 (1984); Genetics Computer Group, University of Wisconsin, Madison, WI), BLASTP, BLASTN, BLASTX, FASTA (Altschul, S.F. et al., J. Molec. Biol. 215:403-410 (1990), PSI-BLAST (Altschul S.F. et al., Nucleic Acids Res. vol. 25, pp. 3389-3402, herein incorporated by reference), eMatrix software (Wu et al., J. Comp. Biol., Vol. 6, pp. 219-235 (1999), herein incorporated by reference), eMotif software (Nevill-Manning et al, ISMB-97, Vol. 4, pp. 202-209, herein incorporated by reference), pFam software (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1), pp. 320-322 (1998), herein incorporated by reference) and the Kyte-Doolittle hydrophobicity prediction algorithm (J. Mol Biol, 157, pp. 105-31 (1982), incorporated herein by reference). The BLAST programs are publicly available from the National Center for Biotechnology Information (NCBI) and other sources (BLAST Manual, Altschul, S., et al. NCB NLM NIH Bethesda, MD 20894; Altschul, S., et al., J. Mol. Biol. 215:403-410 (1990).

#### 4.5 GENE THERAPY

Mutations in the polynucleotides of the invention gene may result in loss of normal function of the encoded protein. The invention thus provides gene therapy to restore normal activity of the polypeptides of the invention; or to treat disease states involving polypeptides of the invention. Delivery of a functional gene encoding polypeptides of the invention to appropriate cells is effected *ex vivo*, *in situ*, or *in vivo* by use of vectors, and more particularly viral vectors (e.g., adenovirus, adeno-associated virus, or a retrovirus), or *ex vivo* by use of physical DNA transfer methods (e.g., liposomes or chemical treatments). See, for example, Anderson, Nature, supplement to vol. 392, no. 6679, pp.25-20 (1998). For additional reviews of gene therapy technology see Friedmann, Science, 244: 1275-1281 (1989); Verma, Scientific American: 68-84 (1990); and Miller, Nature, 357: 455-460 (1992). Introduction of any one of the nucleotides of the present invention or a gene encoding the polypeptides of the present invention can also be accomplished with extrachromosomal substrates (transient

expression) or artificial chromosomes (stable expression). Cells may also be cultured *ex vivo* in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced *in vivo* for therapeutic purposes. Alternatively, it is contemplated that in other human  
5 disease states, preventing the expression of or inhibiting the activity of polypeptides of the invention will be useful in treating the disease states. It is contemplated that antisense therapy or gene therapy could be applied to negatively regulate the expression of polypeptides of the invention.

Other methods inhibiting expression of a protein include the introduction of  
10 antisense molecules to the nucleic acids of the present invention, their complements, or their translated RNA sequences, by methods known in the art. Further, the polypeptides of the present invention can be inhibited by using targeted deletion methods, or the insertion of a negative regulatory element such as a silencer, which is tissue specific.

The present invention still further provides cells genetically engineered *in vivo* to  
15 express the polynucleotides of the invention, wherein such polynucleotides are in operative association with a regulatory sequence heterologous to the host cell which drives expression of the polynucleotides in the cell. These methods can be used to increase or decrease the expression of the polynucleotides of the present invention.

Knowledge of DNA sequences provided by the invention allows for modification of  
20 cells to permit, increase, or decrease, expression of endogenous polypeptide. Cells can be modified (e.g., by homologous recombination) to provide increased polypeptide expression by replacing, in whole or in part, the naturally occurring promoter with all or part of a heterologous promoter so that the cells express the protein at higher levels. The heterologous promoter is inserted in such a manner that it is operatively linked to the desired protein  
25 encoding sequences. See, for example, PCT International Publication No. WO 94/12650, PCT International Publication No. WO 92/20808, and PCT International Publication No. WO 91/09955. It is also contemplated that, in addition to heterologous promoter DNA, amplifiable marker DNA (e.g., *ada*, *dhfr*, and the multifunctional CAD gene which encodes carbamyl phosphate synthase, aspartate transcarbamylase, and dihydroorotase) and/or intron  
30 DNA may be inserted along with the heterologous promoter DNA. If linked to the desired

protein coding sequence, amplification of the marker DNA by standard selection methods results in co-amplification of the desired protein coding sequences in the cells.

In another embodiment of the present invention, cells and tissues may be engineered to express an endogenous gene comprising the polynucleotides of the invention under the control of inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene targeting can be used to replace a gene's existing regulatory region with a regulatory sequence isolated from a different gene or a novel regulatory sequence synthesized by genetic engineering methods. Such regulatory sequences may be comprised of promoters, enhancers, scaffold-attachment regions, negative regulatory elements, transcriptional initiation sites, regulatory protein binding sites or combinations of said sequences. Alternatively, sequences which affect the structure or stability of the RNA or protein produced may be replaced, removed, added, or otherwise modified by targeting. These sequences include polyadenylation signals, mRNA stability elements, splice sites, leader sequences for enhancing or modifying transport or secretion properties of the protein, or other sequences which alter or improve the function or stability of protein or RNA molecules.

The targeting event may be a simple insertion of the regulatory sequence, placing the gene under the control of the new regulatory sequence, *e.g.*, inserting a new promoter or enhancer or both upstream of a gene. Alternatively, the targeting event may be a simple deletion of a regulatory element, such as the deletion of a tissue-specific negative regulatory element. Alternatively, the targeting event may replace an existing element; for example, a tissue-specific enhancer can be replaced by an enhancer that has broader or different cell-type specificity than the naturally occurring elements. Here, the naturally occurring sequences are deleted and new sequences are added. In all cases, the identification of the targeting event may be facilitated by the use of one or more selectable marker genes that are contiguous with the targeting DNA, allowing for the selection of cells in which the exogenous DNA has integrated into the cell genome. The identification of the targeting event may also be facilitated by the use of one or more marker genes exhibiting the property of negative selection, such that the negatively selectable marker is linked to the exogenous DNA, but configured such that the negatively selectable marker flanks the targeting

sequence, and such that a correct homologous recombination event with sequences in the host cell genome does not result in the stable integration of the negatively selectable marker. Markers useful for this purpose include the Herpes Simplex Virus thymidine kinase (TK) gene or the bacterial xanthine-guanine phosphoribosyl-transferase (gpt) gene.

5 The gene targeting or gene activation techniques which can be used in accordance with this aspect of the invention are more particularly described in U.S. Patent No. 5,272,071 to Chappel; U.S. Patent No. 5,578,461 to Sherwin et al.; International Application No. PCT/US92/09627 (WO93/09222) by Selden et al.; and International Application No. PCT/US90/06436 (WO91/06667) by Skoultchi et al., each of which is incorporated by  
10 reference herein in its entirety.

#### 4.6 TRANSGENIC ANIMALS

In preferred methods to determine biological functions of the polypeptides of the invention in vivo, one or more genes provided by the invention are either over expressed  
15 or inactivated in the germ line of animals using homologous recombination [Capecchi, Science 244:1288-1292 (1989)]. Animals in which the gene is over expressed, under the regulatory control of exogenous or endogenous promoter elements, are known as transgenic animals. Animals in which an endogenous gene has been inactivated by homologous recombination are referred to as "knockout" animals. Knockout animals,  
20 preferably non-human mammals, can be prepared as described in U.S. Patent No. 5,557,032, incorporated herein by reference. Transgenic animals are useful to determine the roles polypeptides of the invention play in biological processes, and preferably in disease states. Transgenic animals are useful as model systems to identify compounds that modulate lipid metabolism. Transgenic animals, preferably non-human mammals,  
25 are produced using methods as described in U.S. Patent No 5,489,743 and PCT Publication No. WO94/28122, incorporated herein by reference.

Transgenic animals can be prepared wherein all or part of a promoter of the polynucleotides of the invention is either activated or inactivated to alter the level of expression of the polypeptides of the invention. Inactivation can be carried out using  
30 homologous recombination methods described above. Activation can be achieved by supplementing or even replacing the homologous promoter to provide for increased



protein expression. The homologous promoter can be supplemented by insertion of one or more heterologous enhancer elements known to confer promoter activation in a particular tissue.

The polynucleotides of the present invention also make possible the development, through, e.g., homologous recombination or knock out strategies, of animals that fail to express polypeptides of the invention or that express a variant polypeptide. Such animals are useful as models for studying the *in vivo* activities of polypeptide as well as for studying modulators of the polypeptides of the invention.

In preferred methods to determine biological functions of the polypeptides of the invention *in vivo*, one or more genes provided by the invention are either over expressed or inactivated in the germ line of animals using homologous recombination [Capecchi, Science 244:1288-1292 (1989)]. Animals in which the gene is over expressed, under the regulatory control of exogenous or endogenous promoter elements, are known as transgenic animals. Animals in which an endogenous gene has been inactivated by homologous recombination are referred to as "knockout" animals. Knockout animals, preferably non-human mammals, can be prepared as described in U.S. Patent No. 5,557,032, incorporated herein by reference. Transgenic animals are useful to determine the roles polypeptides of the invention play in biological processes, and preferably in disease states. Transgenic animals are useful as model systems to identify compounds that modulate lipid metabolism. Transgenic animals, preferably non-human mammals, are produced using methods as described in U.S. Patent No 5,489,743 and PCT Publication No. WO94/28122, incorporated herein by reference.

Transgenic animals can be prepared wherein all or part of the polynucleotides of the invention promoter is either activated or inactivated to alter the level of expression of the polypeptides of the invention. Inactivation can be carried out using homologous recombination methods described above. Activation can be achieved by supplementing or even replacing the homologous promoter to provide for increased protein expression. The homologous promoter can be supplemented by insertion of one or more heterologous enhancer elements known to confer promoter activation in a particular tissue.

#### 4.7 USES AND BIOLOGICAL ACTIVITY



labeled) to identify chromosomes or to map related gene positions; to compare with endogenous DNA sequences in patients to identify potential genetic disorders; as probes to hybridize and thus discover novel, related DNA sequences; as a source of information to derive PCR primers for genetic fingerprinting; as a probe to "subtract-out" known sequences in the process of discovering other novel polynucleotides; for selecting and making oligomers for attachment to a "gene chip" or other support, including for examination of expression patterns; to raise anti-protein antibodies using DNA immunization techniques; and as an antigen to raise anti-DNA antibodies or elicit another immune response. Where the polynucleotide encodes a protein which binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the polynucleotide can also be used in interaction trap assays (such as, for example, that described in Gyuris et al., Cell 75:791-803 (1993)) to identify polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction.

The polypeptides provided by the present invention can similarly be used in assays to determine biological activity, including in a panel of multiple proteins for high-throughput screening; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its receptor) in biological fluids; as markers for tissues in which the corresponding polypeptide is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state); and, of course, to isolate correlative receptors or ligands. Proteins involved in these binding interactions can also be used to screen for peptide or small molecule inhibitors or agonists of the binding interaction.

Any or all of these research utilities are capable of being developed into reagent grade or kit format for commercialization as research products.

Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include without limitation "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology:

Guide to Molecular Cloning Techniques", Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987.

#### 4.7.2 NUTRITIONAL USES

Polynucleotides and polypeptides of the present invention can also be used as nutritional sources or supplements. Such uses include without limitation use as a protein or amino acid supplement, use as a carbon source, use as a nitrogen source and use as a source of carbohydrate. In such cases the polypeptide or polynucleotide of the invention can be added to the feed of a particular organism or can be administered as a separate solid or liquid preparation, such as in the form of powder, pills, solutions, suspensions or capsules. In the case of microorganisms, the polypeptide or polynucleotide of the invention can be added to the medium in or on which the microorganism is cultured.

#### 4.7.3 CYTOKINE AND CELL PROLIFERATION/DIFFERENTIATION ACTIVITY

A polypeptide of the present invention may exhibit activity relating to cytokine, cell proliferation (either inducing or inhibiting) or cell differentiation (either inducing or inhibiting) activity or may induce production of other cytokines in certain cell populations. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Many protein factors discovered to date, including all known cytokines, have exhibited activity in one or more factor-dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of therapeutic compositions of the present invention is evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M+(preB M+), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7e, CMK, HUVEC, and Caco. Therapeutic compositions of the invention can be used in the following:

Assays for T-cell or thymocyte proliferation include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function 3.1-3.19;

Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Bertagnolli et al., J. Immunol. 145:1706-1712, 1990; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Bertagnolli, et al., I. Immunol. 149:3778-3783, 1992; Bowman et al., I. Immunol. 152:1756-1761, 1994.

5 Assays for cytokine production and/or proliferation of spleen cells, lymph node cells or thymocytes include, without limitation, those described in: Polyclonal T cell stimulation, Kruisbeek, A. M. and Shevach, E. M. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 3.12.1-3.12.14, John Wiley and Sons, Toronto. 1994; and Measurement of mouse and human interleukin- , Schreiber, R. D. In Current Protocols  
10 in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.8.1-6.8.8, John Wiley and Sons, Toronto. 1994.

Assays for proliferation and differentiation of hematopoietic and lymphopoietic cells include, without limitation, those described in: Measurement of Human and Murine Interleukin 2 and Interleukin 4, Bottomly, K., Davis, L. S. and Lipsky, P. E. In Current  
15 Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.3.1-6.3.12, John Wiley and Sons, Toronto. 1991; deVries et al., J. Exp. Med. 173:1205-1211, 1991; Moreau et al., Nature 336:690-692, 1988; Greenberger et al., Proc. Natl. Acad. Sci. U.S.A. 80:2931-2938, 1983; Measurement of mouse and human interleukin 6--Nordan, R. In  
20 Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.6.1-6.6.5, John Wiley and Sons, Toronto. 1991; Smith et al., Proc. Natl. Acad. Sci. U.S.A. 83:1857-1861, 1986; Measurement of human Interleukin 11--Bennett, F., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.15.1 John  
25 Wiley and Sons, Toronto. 1991; Measurement of mouse and human Interleukin 9--Ciarletta, A., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.13.1, John Wiley and Sons, Toronto. 1991.

Assays for T-cell clone responses to antigens (which will identify, among others, proteins that affect APC-T cell interactions as well as direct T-cell effects by measuring proliferation and cytokine production) include, without limitation, those described in:

Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H.

30 Margulies, E. M. Shevach, W Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function; Chapter

6, Cytokines and their cellular receptors; Chapter 7, Immunologic studies in Humans); Weinberger et al., Proc. Natl. Acad. Sci. USA 77:6091-6095, 1980; Weinberger et al., Eur. J. Immun. 11:405-411, 1981; Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988.

5

#### 4.7.4 STEM CELL GROWTH FACTOR ACTIVITY

A polypeptide of the present invention may exhibit stem cell growth factor activity and be involved in the proliferation, differentiation and survival of pluripotent and totipotent stem cells including primordial germ cells, embryonic stem cells, hematopoietic stem cells and/or germ line stem cells. Administration of the polypeptide of the invention to stem cells *in vivo* or *ex vivo* is expected to maintain and expand cell populations in a totipotent or pluripotent state which would be useful for re-engineering damaged or diseased tissues, transplantation, manufacture of bio-pharmaceuticals and the development of bio-sensors. The ability to produce large quantities of human cells has important working applications for the production of human proteins which currently must be obtained from non-human sources or donors, implantation of cells to treat diseases such as Parkinson's, Alzheimer's and other neurodegenerative diseases; tissues for grafting such as bone marrow, skin, cartilage, tendons, bone, muscle (including cardiac muscle), blood vessels, cornea, neural cells, gastrointestinal cells and others; and organs for transplantation such as kidney, liver, pancreas (including islet cells), heart and lung.

It is contemplated that multiple different exogenous growth factors and/or cytokines may be administered in combination with the polypeptide of the invention to achieve the desired effect, including any of the growth factors listed herein, other stem cell maintenance factors, and specifically including stem cell factor (SCF), leukemia inhibitory factor (LIF), Flt-3 ligand (Flt-3L), any of the interleukins, recombinant soluble IL-6 receptor fused to IL-6, macrophage inflammatory protein 1-alpha (MIP-1-alpha), G-CSF, GM-CSF, thrombopoietin (TPO), platelet factor 4 (PF-4), platelet-derived growth factor (PDGF), neural growth factors and basic fibroblast growth factor (bFGF).

Since totipotent stem cells can give rise to virtually any mature cell type, expansion of these cells in culture will facilitate the production of large quantities of

Stem cells themselves can be transfected with a polynucleotide of the invention to induce autocrine expression of the polypeptide of the invention. This will allow for generation of undifferentiated totipotent/pluripotent stem cell lines that are useful as is or that can then be differentiated into the desired mature cell types. These stable cell lines can also serve as a source of undifferentiated totipotent/pluripotent mRNA to create cDNA libraries and templates for polymerase chain reaction experiments. These studies would allow for the isolation and identification of differentially expressed genes in stem cell populations that regulate stem cell proliferation and/or maintenance.

Expansion and maintenance of totipotent stem cell populations will be useful in the treatment of many pathological conditions. For example, polypeptides of the present invention may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease, accidental damage or genetic disorders. The polypeptide of the invention may be useful for inducing the proliferation of neural cells and for the regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders which involve degeneration, death or trauma to neural cells or nerve tissue. In addition, the expanded stem cell populations can also be genetically altered for gene therapy purposes and to decrease host rejection of replacement tissues after grafting or implantation.

Expression of the polypeptide of the invention and its effect on stem cells can also be manipulated to achieve controlled differentiation of the stem cells into more differentiated cell types. A broadly applicable method of obtaining pure populations of a

specific differentiated cell type from undifferentiated stem cell populations involves the use of a cell-type specific promoter driving a selectable marker. The selectable marker allows only cells of the desired type to survive. For example, stem cells can be induced to differentiate into cardiomyocytes (Wobus et al., Differentiation, 48: 173-182, (1991); Klug et al., J. Clin. Invest., 98(1): 216-224, (1998)) or skeletal muscle cells (Browder, L. W. In: *Principles of Tissue Engineering* eds. Lanza et al., Academic Press (1997)). Alternatively, directed differentiation of stem cells can be accomplished by culturing the stem cells in the presence of a differentiation factor such as retinoic acid and an antagonist of the polypeptide of the invention which would inhibit the effects of endogenous stem cell factor activity and allow differentiation to proceed.

*In vitro* cultures of stem cells can be used to determine if the polypeptide of the invention exhibits stem cell growth factor activity. Stem cells are isolated from any one of various cell sources (including hematopoietic stem cells and embryonic stem cells) and cultured on a feeder layer, as described by Thompson et al. Proc. Natl. Acad. Sci, U.S.A., 92: 7844-7848 (1995), in the presence of the polypeptide of the invention alone or in combination with other growth factors or cytokines. The ability of the polypeptide of the invention to induce stem cells proliferation is determined by colony formation on semi-solid support e.g. as described by Bernstein et al., Blood, 77: 2316-2321 (1991).

#### **4.7.5 HEMATOPOIESIS REGULATING ACTIVITY**

A polypeptide of the present invention may be involved in regulation of hematopoiesis and, consequently, in the treatment of myeloid or lymphoid cell disorders. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates involvement in regulating hematopoiesis, e.g. in supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, thereby indicating utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells; in supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (i.e., traditional CSF activity) useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelo-suppression; in supporting the growth and proliferation



of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of various platelet disorders such as thrombocytopenia, and generally for use in place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells and therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantation, including, without limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either *in-vivo* or *ex-vivo* (i.e., in conjunction with bone marrow transplantation or with peripheral progenitor cell transplantation (homologous or heterologous)) as normal cells or genetically manipulated for gene therapy.

Therapeutic compositions of the invention can be used in the following:

Suitable assays for proliferation and differentiation of various hematopoietic lines are cited above.

Assays for embryonic stem cell differentiation (which will identify, among others, proteins that influence embryonic differentiation hematopoiesis) include, without limitation, those described in: Johansson et al. Cellular Biology 15:141-151, 1995; Keller et al., Molecular and Cellular Biology 13:473-486, 1993; McClanahan et al., Blood 81:2903-2915, 1993.

Assays for stem cell survival and differentiation (which will identify, among others, proteins that regulate lympho-hematopoiesis) include, without limitation, those described in: Methylcellulose colony forming assays, Freshney, M. G. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 265-268, Wiley-Liss, Inc., New York, N.Y. 1994; Hirayama et al., Proc. Natl. Acad. Sci. USA 89:5907-5911, 1992; Primitive hematopoietic colony forming cells with high proliferative potential, McNiece, I. K. and Briddell, R. A. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 23-39, Wiley-Liss, Inc., New York, N.Y. 1994; Neben et al., Experimental Hematology 22:353-359, 1994; Cobblestone area forming cell assay, Ploemacher, R. E. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 1-21, Wiley-Liss, Inc., New York, N.Y. 1994; Long term bone marrow cultures in the presence of stromal cells, Spooncer, E., Dexter, M. and Allen, T. In Culture of Hematopoietic Cells. R. I.

Freshney, et al. eds. Vol pp. 163-179, Wiley-Liss, Inc., New York, N.Y. 1994; Long term culture initiating cell assay, Sutherland, H. J. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 139-162, Wiley-Liss, Inc., New York, N.Y. 1994.

#### 5           **4.7.6   TISSUE GROWTH ACTIVITY**

A polypeptide of the present invention also may be involved in bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration, as well as in wound healing and tissue repair and replacement, and in healing of burns, incisions and ulcers.

10           A polypeptide of the present invention which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Compositions of a polypeptide, antibody, binding partner, or other modulator of the invention may have prophylactic use in closed as well as open fracture reduction and also in the improved fixation of artificial joints. De novo bone formation induced by an  
15           osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

20           A polypeptide of this invention may also be involved in attracting bone-forming cells, stimulating growth of bone-forming cells, or inducing differentiation of progenitors of bone-forming cells. Treatment of osteoporosis, osteoarthritis, bone degenerative disorders, or periodontal disease, such as through stimulation of bone and/or cartilage repair or by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast activity, etc.) mediated by inflammatory processes may also be possible using the composition of the invention.

25           Another category of tissue regeneration activity that may involve the polypeptide of the present invention is tendon/ligament formation. Induction of tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing  
30           damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue.



such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring may allow normal tissue to regenerate. A polypeptide of the present invention may also exhibit angiogenic activity.

A composition of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage.

A composition of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

Therapeutic compositions of the invention can be used in the following:

Assays for tissue generation activity include, without limitation, those described in: International Patent Publication No. WO95/16035 (bone, cartilage, tendon); International Patent Publication No. WO95/05846 (nerve, neuronal); International Patent Publication No. WO91/07491 (skin, endothelium).

Assays for wound healing activity include, without limitation, those described in: Winter, Epidermal Wound Healing, pps. 71-112 (Maibach, H. I. and Rovee, D. T., eds.), Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Mertz, J. Invest. Dermatol 71:382-84 (1978).

#### **4.7.7 IMMUNE STIMULATING OR SUPPRESSING ACTIVITY**

A polypeptide of the present invention may also exhibit immune stimulating or immune suppressing activity, including without limitation the activities for which assays are described herein. A polynucleotide of the invention can encode a polypeptide exhibiting such activities. A protein may be useful in the treatment of various immune deficiencies and disorders (including severe combined immunodeficiency (SCID)), e.g., in regulating (up or down) growth and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by viral (e.g., HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases caused by viral, bacterial, fungal or other infection may be treatable using a protein of the present invention, including infections by HIV, hepatitis viruses, herpes

viruses, mycobacteria, *Leishmania* spp., malaria spp. and various fungal infections such as candidiasis. Of course, in this regard, proteins of the present invention may also be useful where a boost to the immune system generally may be desirable, i.e., in the treatment of cancer.

Autoimmune disorders which may be treated using a protein of the present invention include, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitus, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease.

Such a protein (or antagonists thereof, including antibodies) of the present invention may also to be useful in the treatment of allergic reactions and conditions (*e.g.*, anaphylaxis, serum sickness, drug reactions, food allergies, insect venom allergies, mastocytosis, allergic rhinitis, hypersensitivity pneumonitis, urticaria, angioedema, eczema, atopic dermatitis, allergic contact dermatitis, erythema multiforme, Stevens-Johnson syndrome, allergic conjunctivitis, atopic keratoconjunctivitis, venereal keratoconjunctivitis, giant papillary conjunctivitis and contact allergies), such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for example, organ transplantation), may also be treatable using a protein (or antagonists thereof) of the present invention. The therapeutic effects of the polypeptides or antagonists thereof on allergic reactions can be evaluated by *in vivo* animals models such as the cumulative contact enhancement test (Lastbom et al., Toxicology 125: 59-66, 1998), skin prick test (Hoffmann et al., Allergy 54: 446-54, 1999), guinea pig skin sensitization test (Vohr et al., Arch. Toxicol. 73: 501-9), and murine local lymph node assay (Kimber et al., J. Toxicol. Environ. Health 53: 563-79).

Using the proteins of the invention it may also be possible to modulate immune responses, in a number of ways. Down regulation may be in the form of inhibiting or blocking an immune response already in progress or may involve preventing the induction of an immune response. The functions of activated T cells may be inhibited by suppressing T cell responses or by inducing specific tolerance in T cells, or both.

Immunosuppression of T cell responses is generally an active, non-antigen-specific, process which requires continuous exposure of the T cells to the suppressive agent.

Tolerance, which involves inducing non-responsiveness or anergy in T cells, is distinguishable from immunosuppression in that it is generally antigen-specific and persists after exposure to the tolerizing agent has ceased. Operationally, tolerance can be demonstrated by the lack of a T cell response upon reexposure to specific antigen in the absence of the tolerizing agent.

Down regulating or preventing one or more antigen functions (including without limitation B lymphocyte antigen functions (such as, for example, B7)), e.g., preventing high level lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition as foreign by T cells, followed by an immune reaction that destroys the transplant. The administration of a therapeutic composition of the invention may prevent cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, a lack of costimulation may also be sufficient to anergize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may avoid the necessity of repeated administration of these blocking reagents. To achieve sufficient immunosuppression or tolerance in a subject, it may also be necessary to block the function of a combination of B lymphocyte antigens.

The efficacy of particular therapeutic compositions in preventing organ transplant rejection or GVHD can be assessed using animal models that are predictive of efficacy in humans. Examples of appropriate systems which can be used include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used to examine the immunosuppressive effects of CTLA4Ig fusion proteins in vivo as described in Lenschow et al., Science 257:789-792 (1992) and Turka et al., Proc. Natl. Acad. Sci USA, 89:11102-11105 (1992). In addition, murine models of GVHD (see Paul et al., Fundamental Immunology, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of therapeutic compositions of the invention on the development of that disease.

Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self tissue and which promote the production of cytokines and autoantibodies involved in the pathology of the diseases. Preventing the activation of autoreactive T cells may reduce or eliminate disease symptoms.

Administration of reagents which block stimulation of T cells can be used to inhibit T cell activation and prevent production of autoantibodies or T cell-derived cytokines which may be involved in the disease process. Additionally, blocking reagents may induce antigen-specific tolerance of autoreactive T cells which could lead to long-term relief from the disease. The efficacy of blocking reagents in preventing or alleviating autoimmune disorders can be determined using a number of well-characterized animal models of human autoimmune diseases. Examples include murine experimental autoimmune encephalitis, systemic lupus erythmatosis in MRL/lpr/lpr mice or NZB hybrid mice, murine autoimmune collagen arthritis, diabetes mellitus in NOD mice and BB rats, and murine experimental myasthenia gravis (see Paul ed., *Fundamental Immunology*, Raven Press, New York, 1989, pp. 840-856).

Upregulation of an antigen function (e.g., a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy. Upregulation of immune responses may be in the form of enhancing an existing immune response or eliciting an initial immune response. For example, enhancing an immune response may be useful in cases of viral infection, including systemic viral diseases such as influenza, the common cold, and encephalitis.

Alternatively, anti-viral immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells in vitro with viral antigen-pulsed APCs either expressing a peptide of the present invention or together with a stimulatory form of a soluble peptide of the present invention and reintroducing the in vitro activated T cells into the patient. Another method of enhancing anti-viral immune responses would be to isolate infected cells from a patient, transfect them with a nucleic acid encoding a protein of the present invention as described herein such that the cells express all or a portion of the protein on their surface, and reintroduce the transfected

cells into the patient. The infected cells would now be capable of delivering a costimulatory signal to, and thereby activate, T cells in vivo.

A polypeptide of the present invention may provide the necessary stimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack MHC class I or MHC class II molecules, or which fail to reexpress sufficient mounts of MHC class I or MHC class II molecules, can be transfected with nucleic acid encoding all or a portion of (e.g., a cytoplasmic-domain truncated portion) of an MHC class I alpha chain protein and  $\beta_2$  microglobulin protein or an MHC class II alpha chain protein and an MHC class II beta chain protein to thereby express MHC class I or MHC class II proteins on the cell surface. Expression of the appropriate class I or class II MHC in conjunction with a peptide having the activity of a B lymphocyte antigen (e.g., B7-1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can also be cotransfected with a DNA encoding a peptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for thymocyte or splenocyte cytotoxicity include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Herrmann et al., Proc. Natl. Acad. Sci. USA 78:2488-2492, 1981; Herrmann et al., J. Immunol. 128:1968-1974, 1982; Handa et al., J. Immunol. 135:1564-1572, 1985; Takai et al., I. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bowman et al., J. Virology 61:1992-1998; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Brown et al., J. Immunol. 153:3079-3092, 1994.



Assays for T-cell-dependent immunoglobulin responses and isotype switching (which will identify, among others, proteins that modulate T-cell dependent antibody responses and that affect Th1/Th2 profiles) include, without limitation, those described in: Maliszewski, J. Immunol. 144:3028-3033, 1990; and Assays for B cell function: In vitro antibody production, Mond, J. J. and Brunswick, M. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 3.8.1-3.8.16, John Wiley and Sons, Toronto. 1994.

Mixed lymphocyte reaction (MLR) assays (which will identify, among others, proteins that generate predominantly Th1 and CTL responses) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bertagnolli et al., J. Immunol. 149:3778-3783, 1992.

Dendritic cell-dependent assays (which will identify, among others, proteins expressed by dendritic cells that activate naive T-cells) include, without limitation, those described in: Guery et al., J. Immunol. 134:536-544, 1995; Inaba et al., Journal of Experimental Medicine 173:549-559, 1991; Macatonia et al., Journal of Immunology 154:5071-5079, 1995; Porgador et al., Journal of Experimental Medicine 182:255-260, 1995; Nair et al., Journal of Virology 67:4062-4069, 1993; Huang et al., Science 264:961-965, 1994; Macatonia et al., Journal of Experimental Medicine 169:1255-1264, 1989; Bhardwaj et al., Journal of Clinical Investigation 94:797-807, 1994; and Inaba et al., Journal of Experimental Medicine 172:631-640, 1990.

Assays for lymphocyte survival/apoptosis (which will identify, among others, proteins that prevent apoptosis after superantigen induction and proteins that regulate lymphocyte homeostasis) include, without limitation, those described in: Darzynkiewicz et al., Cytometry 13:795-808, 1992; Gorczyca et al., Leukemia 7:659-670, 1993; Gorczyca et al., Cancer Research 53:1945-1951, 1993; Itoh et al., Cell 66:233-243, 1991; Zacharchuk, Journal of Immunology 145:4037-4045, 1990; Zamai et al., Cytometry 14:891-897, 1993; Gorczyca et al., International Journal of Oncology 1:639-648, 1992.

Assays for proteins that influence early steps of T-cell commitment and development include, without limitation, those described in: Antica et al., Blood 84:111-117, 1994; Fine et al., Cellular Immunology 155:111-122, 1994; Galy et al., Blood 85:2770-2778, 1995; Toki et al., Proc. Nat. Acad Sci. USA 88:7548-7551, 1991.

#### 4.7.8 ACTIVIN/INHIBIN ACTIVITY

A polypeptide of the present invention may also exhibit activin- or inhibin-related activities. A polynucleotide of the invention may encode a polypeptide exhibiting such characteristics. Inhibins are characterized by their ability to inhibit the release of follicle stimulating hormone (FSH), while activins are characterized by their ability to stimulate the release of follicle stimulating hormone (FSH). Thus, a polypeptide of the present invention, alone or in heterodimers with a member of the inhibin family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the polypeptide of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin group, may be useful as a fertility inducing therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example, U.S. Pat. No. 4,798,885. A polypeptide of the invention may also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as, but not limited to, cows, sheep and pigs.

The activity of a polypeptide of the invention may, among other means, be measured by the following methods.

Assays for activin/inhibin activity include, without limitation, those described in: Vale et al., Endocrinology 91:562-572, 1972; Ling et al., Nature 321:779-782, 1986; Vale et al., Nature 321:776-779, 1986; Mason et al., Nature 318:659-663, 1985; Forage et al., Proc. Natl. Acad. Sci. USA 83:3091-3095, 1986.

#### 4.7.9 CHEMOTACTIC/CHEMOKINETIC ACTIVITY

A polypeptide of the present invention may be involved in chemotactic or chemokinetic activity for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Chemotactic and chemokinetic receptor activation can be used to mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic compositions (e.g. proteins, antibodies, binding partners, or modulators of the invention) provide particular advantages in treatment of wounds and other trauma to tissues, as well as in treatment of localized infections. For example, attraction of lymphocytes, monocytes or neutrophils to tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell population. Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

Therapeutic compositions of the invention can be used in the following:

Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein to induce the adhesion of one cell population to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Marguiles, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 6.12, Measurement of alpha and beta Chemokines 6.12.1-6.12.28; Taub et al. J. Clin. Invest. 95:1370-1376, 1995; Lind et al. APMIS 103:140-146, 1995; Muller et al Eur. J. Immunol. 25:1744-1748; Gruber et al. J. of Immunol. 152:5860-5867, 1994; Johnston et al. J. of Immunol. 153:1762-1768, 1994.

#### 4.7.10 HEMOSTATIC AND THROMBOLYTIC ACTIVITY

A polypeptide of the invention may also be involved in hemostasis or thrombolysis or thrombosis. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Compositions may be useful in treatment of various coagulation disorders (including hereditary disorders, such as hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A composition of the invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom (such as, for example, infarction of cardiac and central nervous system vessels (e.g., stroke).

Therapeutic compositions of the invention can be used in the following:

Assays for hemostatic and thrombolytic activity include, without limitation, those described in: Linet et al., J. Clin. Pharmacol. 26:131-140, 1986; Burdick et al., Thrombosis Res. 45:413-419, 1987; Humphrey et al., Fibrinolysis 5:71-79 (1991); Schaub, Prostaglandins 35:467-474, 1988.

#### **4.7.11 CANCER DIAGNOSIS AND THERAPY**

Polypeptides of the invention may be involved in cancer cell generation, proliferation or metastasis. Detection of the presence or amount of polynucleotides or polypeptides of the invention may be useful for the diagnosis and/or prognosis of one or more types of cancer. For example, the presence or increased expression of a polynucleotide/polypeptide of the invention may indicate a hereditary risk of cancer, a precancerous condition, or an ongoing malignancy. Conversely, a defect in the gene or absence of the polypeptide may be associated with a cancer condition. Identification of single nucleotide polymorphisms associated with cancer or a predisposition to cancer may also be useful for diagnosis or prognosis.

Cancer treatments promote tumor regression by inhibiting tumor cell proliferation, inhibiting angiogenesis (growth of new blood vessels that is necessary to support tumor growth) and/or prohibiting metastasis by reducing tumor cell motility or invasiveness. Therapeutic compositions of the invention may be effective in adult and pediatric oncology including in solid phase tumors/malignancies, locally advanced tumors, human soft tissue sarcomas, metastatic cancer, including lymphatic metastases,



Doxorubicin HCl, Estramustine phosphate sodium, Etoposide (V16-213), Floxuridine, 5-Fluorouracil (5-Fu), Flutamide, Hydroxyurea (hydroxycarbamide), Ifosfamide, Interferon Alpha-2a, Interferon Alpha-2b, Leuprolide acetate (LHRH-releasing factor analog), Lomustine, Mechlorethamine HCl (nitrogen mustard), Melphalan, Mercaptopurine, Mesna, Methotrexate (MTX), Mitomycin, Mitoxantrone HCl, Octreotide, Plicamycin, Procarbazine HCl, Streptozocin, Tamoxifen citrate, Thioguanine, Thiotepa, Vinblastine sulfate, Vincristine sulfate, Amsacrine, Azacitidine, Hexamethylmelamine, Interleukin-2, Mitoguanzone, Pentostatin, Semustine, Teniposide, and Vindesine sulfate.

In addition, therapeutic compositions of the invention may be used for prophylactic treatment of cancer. There are hereditary conditions and/or environmental situations (e.g. exposure to carcinogens) known in the art that predispose an individual to developing cancers. Under these circumstances, it may be beneficial to treat these individuals with therapeutically effective doses of the polypeptide of the invention to reduce the risk of developing cancers.

*In vitro* models can be used to determine the effective doses of the polypeptide of the invention as a potential cancer treatment. These *in vitro* models include proliferation assays of cultured tumor cells, growth of cultured tumor cells in soft agar (see Freshney, (1987) Culture of Animal Cells: A Manual of Basic Technique, Wiley-Liss, New York, NY Ch 18 and Ch 21), tumor systems in nude mice as described in Giovanella et al., J. Natl. Can. Inst., 52: 921-30 (1974), mobility and invasive potential of tumor cells in Boyden Chamber assays as described in Pilkington et al., Anticancer Res., 17: 4107-9 (1997), and angiogenesis assays such as induction of vascularization of the chick chorioallantoic membrane or induction of vascular endothelial cell migration as described in Ribatta et al., Intl. J. Dev. Biol., 40: 1189-97 (1999) and Li et al., Clin. Exp. Metastasis, 17:423-9 (1999), respectively. Suitable tumor cells lines are available, e.g. from American Type Tissue Culture Collection catalogs.

#### 4.7.12 RECEPTOR/LIGAND ACTIVITY

A polypeptide of the present invention may also demonstrate activity as receptor, receptor ligand or inhibitor or agonist of receptor/ligand interactions. A polynucleotide of the invention can encode a polypeptide exhibiting such characteristics. Examples of

such receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and humoral immune responses. Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions.

The activity of a polypeptide of the invention may, among other means, be measured by the following methods:

Suitable assays for receptor-ligand activity include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley- Interscience (Chapter 7.28, Measurement of Cellular Adhesion under static conditions 7.28.1- 7.28.22), Takai et al., Proc. Natl. Acad. Sci. USA 84:6864-6868, 1987; Bierer et al., J. Exp. Med. 168:1145-1156, 1988; Rosenstein et al., J. Exp. Med. 169:149-160 1989; Stoltenborg et al., J. Immunol. Methods 175:59-68, 1994; Stitt et al., Cell 80:661-670, 1995.

By way of example, the polypeptides of the invention may be used as a receptor for a ligand(s) thereby transmitting the biological activity of that ligand(s). Ligands may be identified through binding assays, affinity chromatography, dihybrid screening assays, BIAcore assays, gel overlay assays, or other methods known in the art.

Studies characterizing drugs or proteins as agonist or antagonist or partial agonists or a partial antagonist require the use of other proteins as competing ligands. The polypeptides of the present invention or ligand(s) thereof may be labeled by being coupled to radioisotopes, colorimetric molecules or a toxin molecules by conventional methods. ("Guide to Protein Purification" Murray P. Deutscher (ed) Methods in Enzymology Vol. 182 (1990) Academic Press, Inc. San Diego). Examples of radioisotopes include, but are not limited to, tritium and carbon-14 . Examples of

colorimetric molecules include, but are not limited to, fluorescent molecules such as fluorescamine, or rhodamine or other colorimetric molecules. Examples of toxins include, but are not limited, to ricin.

#### 5           **4.7.13 DRUG SCREENING**

10           This invention is particularly useful for screening chemical compounds by using the novel polypeptides or binding fragments thereof in any of a variety of drug screening techniques. The polypeptides or fragments employed in such a test may either be free in solution, affixed to a solid support, borne on a cell surface or located intracellularly. One method of drug screening utilizes eukaryotic or prokaryotic host cells which are stably transformed with recombinant nucleic acids expressing the polypeptide or a fragment thereof. Drugs are screened against such transformed cells in competitive binding assays. Such cells, either in viable or fixed form, can be used for standard binding assays. One may measure, for example, the formation of complexes between polypeptides of the invention or fragments and the agent being tested or examine the diminution in complex formation between the novel polypeptides and an appropriate cell line, which are well known in the art.

20           Sources for test compounds that may be screened for ability to bind to or modulate (i.e., increase or decrease) the activity of polypeptides of the invention include (1) inorganic and organic chemical libraries, (2) natural product libraries, and (3) combinatorial libraries comprised of either random or mimetic peptides, oligonucleotides or organic molecules.

25           Chemical libraries may be readily synthesized or purchased from a number of commercial sources, and may include structural analogs of known compounds or compounds that are identified as "hits" or "leads" via natural product screening.

30           The sources of natural product libraries are microorganisms (including bacteria and fungi), animals, plants or other vegetation, or marine organisms, and libraries of mixtures for screening may be created by: (1) fermentation and extraction of broths from soil, plant or marine microorganisms or (2) extraction of the organisms themselves. Natural product libraries include polyketides, non-ribosomal peptides, and (non-naturally occurring) variants thereof. For a review, see *Science* 282:63-68 (1998).



Combinatorial libraries are composed of large numbers of peptides, oligonucleotides or organic compounds and can be readily prepared by traditional automated synthesis methods, PCR, cloning or proprietary synthetic methods. Of particular interest are peptide and oligonucleotide combinatorial libraries. Still other libraries of interest include peptide, protein, peptidomimetic, multiparallel synthetic collection, recombinatorial, and polypeptide libraries. For a review of combinatorial chemistry and libraries created therefrom, see Myers, *Curr. Opin. Biotechnol.* 8:701-707 (1997). For reviews and examples of peptidomimetic libraries, see Al-Obeidi et al., *Mol. Biotechnol.*, 9(3):205-23 (1998); Hruby et al., *Curr Opin Chem Biol*, 1(1):114-19 (1997); Dorner et al., *Bioorg Med Chem*, 4(5):709-15 (1996) (alkylated dipeptides).

Identification of modulators through use of the various libraries described herein permits modification of the candidate "hit" (or "lead") to optimize the capacity of the "hit" to bind a polypeptide of the invention. The molecules identified in the binding assay are then tested for antagonist or agonist activity in *in vivo* tissue culture or animal models that are well known in the art. In brief, the molecules are titrated into a plurality of cell cultures or animals and then tested for either cell/animal death or prolonged survival of the animal/cells.

The binding molecules thus identified may be complexed with toxins, e.g., ricin or cholera, or with other compounds that are toxic to cells such as radioisotopes. The toxin-binding molecule complex is then targeted to a tumor or other cell by the specificity of the binding molecule for a polypeptide of the invention. Alternatively, the binding molecules may be complexed with imaging agents for targeting and imaging purposes.

#### 4.7.14 ASSAY FOR RECEPTOR ACTIVITY

The invention also provides methods to detect specific binding of a polypeptide e.g. a ligand or a receptor. The art provides numerous assays particularly useful for identifying previously unknown binding partners for receptor polypeptides of the invention. For example, expression cloning using mammalian or bacterial cells, or dihybrid screening assays can be used to identify polynucleotides encoding binding partners. As another example, affinity chromatography with the appropriate immobilized polypeptide of the invention can be used to isolate polypeptides that recognize and bind

polypeptides of the invention. There are a number of different libraries used for the identification of compounds, and in particular small molecules, that modulate (*i.e.*, increase or decrease) biological activity of a polypeptide of the invention. Ligands for receptor polypeptides of the invention can also be identified by adding exogenous ligands, or cocktails of ligands to two cells populations that are genetically identical except for the expression of the receptor of the invention: one cell population expresses the receptor of the invention whereas the other does not. The response of the two cell populations to the addition of ligands(s) are then compared. Alternatively, an expression library can be co-expressed with the polypeptide of the invention in cells and assayed for an autocrine response to identify potential ligand(s). As still another example, BIAcore assays, gel overlay assays, or other methods known in the art can be used to identify binding partner polypeptides, including, (1) organic and inorganic chemical libraries, (2) natural product libraries, and (3) combinatorial libraries comprised of random peptides, oligonucleotides or organic molecules.

The role of downstream intracellular signaling molecules in the signaling cascade of the polypeptide of the invention can be determined. For example, a chimeric protein in which the cytoplasmic domain of the polypeptide of the invention is fused to the extracellular portion of a protein, whose ligand has been identified, is produced in a host cell. The cell is then incubated with the ligand specific for the extracellular portion of the chimeric protein, thereby activating the chimeric receptor. Known downstream proteins involved in intracellular signaling can then be assayed for expected modifications *i.e.* phosphorylation. Other methods known to those in the art can also be used to identify signaling molecules involved in receptor activity.

#### **4.7.15 ANTI-INFLAMMATORY ACTIVITY**

Compositions of the present invention may also exhibit anti-inflammatory activity. The anti-inflammatory activity may be achieved by providing a stimulus to cells involved in the inflammatory response, by inhibiting or promoting cell-cell interactions (such as, for example, cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production of other factors which more directly inhibit or

promote an inflammatory response. Compositions with such activities can be used to treat inflammatory conditions including chronic or acute conditions), including without limitation intimation associated with infection (such as septic shock, sepsis or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine-induced lung injury, inflammatory bowel disease, Crohn's disease or resulting from over production of cytokines such as TNF or IL-1. Compositions of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material. Compositions of this invention may be utilized to prevent or treat conditions such as, but not limited to, sepsis, acute pancreatitis, endotoxin shock, cytokine induced shock, rheumatoid arthritis, chronic inflammatory arthritis, pancreatic cell damage from diabetes mellitus type 1, graft versus host disease, inflammatory bowel disease, inflammation associated with pulmonary disease, other autoimmune disease or inflammatory disease, an antiproliferative agent such as for acute or chronic myelogenous leukemia or in the prevention of premature labor secondary to intrauterine infections.

#### **4.7.16 LEUKEMIAS**

Leukemias and related disorders may be treated or prevented by administration of a therapeutic that promotes or inhibits function of the polynucleotides and/or polypeptides of the invention. Such leukemias and related disorders include but are not limited to acute leukemia, acute lymphocytic leukemia, acute myelocytic leukemia, myeloblastic, promyelocytic, myelomonocytic, monocytic, erythroleukemia, chronic leukemia, chronic myelocytic (granulocytic) leukemia and chronic lymphocytic leukemia (for a review of such disorders, see Fishman et al., 1985, Medicine, 2d Ed., J.B. Lippincott Co., Philadelphia).

#### **4.7.17 NERVOUS SYSTEM DISORDERS**

Nervous system disorders, involving cell types which can be tested for efficacy of intervention with compounds that modulate the activity of the polynucleotides and/or polypeptides of the invention, and which can be treated upon thus observing an indication of therapeutic utility, include but are not limited to nervous system injuries, and diseases

or disorders which result in either a disconnection of axons, a diminution or degeneration of neurons, or demyelination. Nervous system lesions which may be treated in a patient (including human and non-human mammalian patients) according to the invention include but are not limited to the following lesions of either the central (including spinal cord, brain) or peripheral nervous systems:

(i) traumatic lesions, including lesions caused by physical injury or associated with surgery, for example, lesions which sever a portion of the nervous system, or compression injuries;

(ii) ischemic lesions, in which a lack of oxygen in a portion of the nervous system results in neuronal injury or death, including cerebral infarction or ischemia, or spinal cord infarction or ischemia;

(iii) infectious lesions, in which a portion of the nervous system is destroyed or injured as a result of infection, for example, by an abscess or associated with infection by human immunodeficiency virus, herpes zoster, or herpes simplex virus or with Lyme disease, tuberculosis, syphilis;

(iv) degenerative lesions, in which a portion of the nervous system is destroyed or injured as a result of a degenerative process including but not limited to degeneration associated with Parkinson's disease, Alzheimer's disease, Huntington's chorea, or amyotrophic lateral sclerosis;

(v) lesions associated with nutritional diseases or disorders, in which a portion of the nervous system is destroyed or injured by a nutritional disorder or disorder of metabolism including but not limited to, vitamin B12 deficiency, folic acid deficiency, Wernicke disease, tobacco-alcohol amblyopia, Marchiafava-Bignami disease (primary degeneration of the corpus callosum), and alcoholic cerebellar degeneration;

(vi) neurological lesions associated with systemic diseases including but not limited to diabetes (diabetic neuropathy, Bell's palsy), systemic lupus erythematosus, carcinoma, or sarcoidosis;

(vii) lesions caused by toxic substances including alcohol, lead, or particular neurotoxins; and

(viii) demyelinated lesions in which a portion of the nervous system is destroyed or injured by a demyelinating disease including but not limited to multiple

sclerosis, human immunodeficiency virus-associated myelopathy, transverse myelopathy or various etiologies, progressive multifocal leukoencephalopathy, and central pontine myelinolysis.

Therapeutics which are useful according to the invention for treatment of a nervous system disorder may be selected by testing for biological activity in promoting the survival or differentiation of neurons. For example, and not by way of limitation, therapeutics which elicit any of the following effects may be useful according to the invention:

- (i) increased survival time of neurons in culture;
- (ii) increased sprouting of neurons in culture or *in vivo*;
- (iii) increased production of a neuron-associated molecule in culture or *in vivo*, *e.g.*, choline acetyltransferase or acetylcholinesterase with respect to motor neurons; or
- (iv) decreased symptoms of neuron dysfunction *in vivo*.

Such effects may be measured by any method known in the art. In preferred, non-limiting embodiments, increased survival of neurons may be measured by the method set forth in Arakawa et al. (1990, J. Neurosci. 10:3507-3515); increased sprouting of neurons may be detected by methods set forth in Pestronk et al. (1980, Exp. Neurol. 70:65-82) or Brown et al. (1981, Ann. Rev. Neurosci. 4:17-42); increased production of neuron-associated molecules may be measured by bioassay, enzymatic assay, antibody binding, Northern blot assay, *etc.*, depending on the molecule to be measured; and motor neuron dysfunction may be measured by assessing the physical manifestation of motor neuron disorder, *e.g.*, weakness, motor neuron conduction velocity, or functional disability.

In specific embodiments, motor neuron disorders that may be treated according to the invention include but are not limited to disorders such as infarction, infection, exposure to toxin, trauma, surgical damage, degenerative disease or malignancy that may affect motor neurons as well as other components of the nervous system, as well as disorders that selectively affect neurons such as amyotrophic lateral sclerosis, and including but not limited to progressive spinal muscular atrophy, progressive bulbar palsy, primary lateral sclerosis, infantile and juvenile muscular atrophy, progressive

bulbar paralysis of childhood (Fazio-Londe syndrome), poliomyelitis and the post polio syndrome, and Hereditary Motorsensory Neuropathy (Charcot-Marie-Tooth Disease).

#### 4.7.18 OTHER ACTIVITIES

5 A polypeptide of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing, infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites; effecting (suppressing or enhancing) bodily characteristics, including, without limitation, height, weight, hair color, eye color, skin, fat to lean ratio or other tissue  
10 pigmentation, or organ or body part size or shape (such as, for example, breast augmentation or diminution, change in bone form or shape); effecting biorhythms or circadian cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, co-factors or other nutritional  
15 factors or component(s); effecting behavioral characteristics, including, without limitation, appetite, libido, stress, cognition (including cognitive disorders), depression (including depressive disorders) and violent behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case  
20 of enzymes, correcting deficiencies of the enzyme and treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in a vaccine composition to raise an immune response against such protein or another material or entity which is  
25 cross-reactive with such protein.

#### 4.7.19 IDENTIFICATION OF POLYMORPHISMS

The demonstration of polymorphisms makes possible the identification of such polymorphisms in human subjects and the pharmacogenetic use of this information for  
30 diagnosis and treatment. Such polymorphisms may be associated with, e.g., differential predisposition or susceptibility to various disease states (such as disorders involving

inflammation or immune response) or a differential response to drug administration, and this genetic information can be used to tailor preventive or therapeutic treatment appropriately. For example, the existence of a polymorphism associated with a predisposition to inflammation or autoimmune disease makes possible the diagnosis of this condition in humans by identifying the presence of the polymorphism.

Polymorphisms can be identified in a variety of ways known in the art which all generally involve obtaining a sample from a patient, analyzing DNA from the sample, optionally involving isolation or amplification of the DNA, and identifying the presence of the polymorphism in the DNA. For example, PCR may be used to amplify an appropriate fragment of genomic DNA which may then be sequenced. Alternatively, the DNA may be subjected to allele-specific oligonucleotide hybridization (in which appropriate oligonucleotides are hybridized to the DNA under conditions permitting detection of a single base mismatch) or to a single nucleotide extension assay (in which an oligonucleotide that hybridizes immediately adjacent to the position of the polymorphism is extended with one or more labeled nucleotides). In addition, traditional restriction fragment length polymorphism analysis (using restriction enzymes that provide differential digestion of the genomic DNA depending on the presence or absence of the polymorphism) may be performed. Arrays with nucleotide sequences of the present invention can be used to detect polymorphisms. The array can comprise modified nucleotide sequences of the present invention in order to detect the nucleotide sequences of the present invention. In the alternative, any one of the nucleotide sequences of the present invention can be placed on the array to detect changes from those sequences.

Alternatively a polymorphism resulting in a change in the amino acid sequence could also be detected by detecting a corresponding change in amino acid sequence of the protein, e.g., by an antibody specific to the variant sequence.

#### **4.7.20 ARTHRITIS AND INFLAMMATION**

The immunosuppressive effects of the compositions of the invention against rheumatoid arthritis is determined in an experimental animal model system. The experimental model system is adjuvant induced arthritis in rats, and the protocol is described by J. Holoshitz, et al., 1983, Science, 219:56, or by B. Waksman et al., 1963,

Int. Arch. Allergy Appl. Immunol., 23:129. Induction of the disease can be caused by a single injection, generally intradermally, of a suspension of killed Mycobacterium tuberculosis in complete Freund's adjuvant (CFA). The route of injection can vary, but rats may be injected at the base of the tail with an adjuvant mixture. The polypeptide is administered in phosphate buffered solution (PBS) at a dose of about 1-5 mg/kg. The control consists of administering PBS only.

The procedure for testing the effects of the test compound would consist of intradermally injecting killed Mycobacterium tuberculosis in CFA followed by immediately administering the test compound and subsequent treatment every other day until day 24. At 14, 15, 18, 20, 22, and 24 days after injection of Mycobacterium CFA, an overall arthritis score may be obtained as described by J. Holoskitz above. An analysis of the data would reveal that the test compound would have a dramatic affect on the swelling of the joints as measured by a decrease of the arthritis score.

#### 4.8 THERAPEUTIC METHODS

The compositions (including polypeptide fragments, analogs, variants and antibodies or other binding partners or modulators including antisense polynucleotides) of the invention have numerous applications in a variety of therapeutic methods.

Examples of therapeutic applications include, but are not limited to, those exemplified herein.

##### 4.8.1 EXAMPLE

One embodiment of the invention is the administration of an effective amount of the polypeptides or other composition of the invention to individuals affected by a disease or disorder that can be modulated by regulating the peptides of the invention. While the mode of administration is not particularly important, parenteral administration is preferred. An exemplary mode of administration is to deliver an intravenous bolus. The dosage of the polypeptides or other composition of the invention will normally be determined by the prescribing physician. It is to be expected that the dosage will vary according to the age, weight, condition and response of the individual patient. Typically,



the amount of polypeptide administered per dose will be in the range of about 0.01µg/kg to 100 mg/kg of body weight, with the preferred dose being about 0.1µg/kg to 10 mg/kg of patient body weight. For parenteral administration, polypeptides of the invention will be formulated in an injectable form combined with a pharmaceutically acceptable parenteral vehicle. Such vehicles are well known in the art and examples include water, saline, Ringer's solution, dextrose solution, and solutions consisting of small amounts of the human serum albumin. The vehicle may contain minor amounts of additives that maintain the isotonicity and stability of the polypeptide or other active ingredient. The preparation of such solutions is within the skill of the art.

#### **4.9 PHARMACEUTICAL FORMULATIONS AND ROUTES OF ADMINISTRATION**

A protein or other composition of the present invention (from whatever source derived, including without limitation from recombinant and non-recombinant sources and including antibodies and other binding partners of the polypeptides of the invention) may be administered to a patient in need, by itself, or in pharmaceutical compositions where it is mixed with suitable carriers or excipient(s) at doses to treat or ameliorate a variety of disorders. Such a composition may optionally contain (in addition to protein or other active ingredient and a carrier) diluents, fillers, salts, buffers, stabilizers, solubilizers, and other materials well known in the art. The term "pharmaceutically acceptable" means a non-toxic material that does not interfere with the effectiveness of the biological activity of the active ingredient(s). The characteristics of the carrier will depend on the route of administration. The pharmaceutical composition of the invention may also contain cytokines, lymphokines, or other hematopoietic factors such as M-CSF, GM-CSF, TNF, IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IFN, TNF0, TNF1, TNF2, G-CSF, Meg-CSF, thrombopoietin, stem cell factor, and erythropoietin. In further compositions, proteins of the invention may be combined with other agents beneficial to the treatment of the disease or disorder in question. These agents include various growth factors such as epidermal growth factor (EGF),

platelet-derived growth factor (PDGF), transforming growth factors (TGF- $\alpha$  and TGF- $\beta$ ), insulin-like growth factor (IGF), as well as cytokines described herein.

The pharmaceutical composition may further contain other agents which either enhance the activity of the protein or other active ingredient or complement its activity or use in treatment. Such additional factors and/or agents may be included in the pharmaceutical composition to produce a synergistic effect with protein or other active ingredient of the invention, or to minimize side effects. Conversely, protein or other active ingredient of the present invention may be included in formulations of the particular clotting factor, cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti-inflammatory agent to minimize side effects of the clotting factor, cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti-inflammatory agent (such as IL-1Ra, IL-1 Hy1, IL-1 Hy2, anti-TNF, corticosteroids, immunosuppressive agents). A protein of the present invention may be active in multimers (e.g., heterodimers or homodimers) or complexes with itself or other proteins. As a result, pharmaceutical compositions of the invention may comprise a protein of the invention in such multimeric or complexed form.

As an alternative to being included in a pharmaceutical composition of the invention including a first protein, a second protein or a therapeutic agent may be concurrently administered with the first protein (e.g., at the same time, or at differing times provided that therapeutic concentrations of the combination of agents is achieved at the treatment site). Techniques for formulation and administration of the compounds of the instant application may be found in "Remington's Pharmaceutical Sciences," Mack Publishing Co., Easton, PA, latest edition. A therapeutically effective dose further refers to that amount of the compound sufficient to result in amelioration of symptoms, e.g., treatment, healing, prevention or amelioration of the relevant medical condition, or an increase in rate of treatment, healing, prevention or amelioration of such conditions. When applied to an individual active ingredient, administered alone, a therapeutically effective dose refers to that ingredient alone. When applied to a combination, a therapeutically effective dose refers to combined amounts of the active ingredients that result in the therapeutic effect, whether administered in combination, serially or simultaneously.

In practicing the method of treatment or use of the present invention, a therapeutically effective amount of protein or other active ingredient of the present invention is administered to a mammal having a condition to be treated. Protein or other active ingredient of the present invention may be administered in accordance with the method of the invention either alone or in combination with other therapies such as treatments employing cytokines, lymphokines or other hematopoietic factors. When co-administered with one or more cytokines, lymphokines or other hematopoietic factors, protein or other active ingredient of the present invention may be administered either simultaneously with the cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors, or sequentially. If administered sequentially, the attending physician will decide on the appropriate sequence of administering protein or other active ingredient of the present invention in combination with cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors.

#### **4.9.1 ROUTES OF ADMINISTRATION**

Suitable routes of administration may, for example, include oral, rectal, transmucosal, or intestinal administration; parenteral delivery, including intramuscular, subcutaneous, intramedullary injections, as well as intrathecal, direct intraventricular, intravenous, intraperitoneal, intranasal, or intraocular injections. Administration of protein or other active ingredient of the present invention used in the pharmaceutical composition or to practice the method of the present invention can be carried out in a variety of conventional ways, such as oral ingestion, inhalation, topical application or cutaneous, subcutaneous, intraperitoneal, parenteral or intravenous injection. Intravenous administration to the patient is preferred.

Alternately, one may administer the compound in a local rather than systemic manner, for example, via injection of the compound directly into a arthritic joints or in fibrotic tissue, often in a depot or sustained release formulation. In order to prevent the scarring process frequently occurring as complication of glaucoma surgery, the compounds may be administered topically, for example, as eye drops. Furthermore, one may administer the drug in a targeted drug delivery system, for example, in a liposome

coated with a specific antibody, targeting, for example, arthritic or fibrotic tissue. The liposomes will be targeted to and taken up selectively by the afflicted tissue.

The polypeptides of the invention are administered by any route that delivers an effective dosage to the desired site of action. The determination of a suitable route of administration and an effective dosage for a particular indication is within the level of skill in the art. Preferably for wound treatment, one administers the therapeutic compound directly to the site. Suitable dosage ranges for the polypeptides of the invention can be extrapolated from these dosages or from similar studies in appropriate animal models. Dosages can then be adjusted as necessary by the clinician to provide maximal therapeutic benefit.

#### 4.9.2 COMPOSITIONS/FORMULATIONS

Pharmaceutical compositions for use in accordance with the present invention thus may be formulated in a conventional manner using one or more physiologically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. These pharmaceutical compositions may be manufactured in a manner that is itself known, *e.g.*, by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes. Proper formulation is dependent upon the route of administration chosen. When a therapeutically effective amount of protein or other active ingredient of the present invention is administered orally, protein or other active ingredient of the present invention will be in the form of a tablet, capsule, powder, solution or elixir. When administered in tablet form, the pharmaceutical composition of the invention may additionally contain a solid carrier such as a gelatin or an adjuvant. The tablet, capsule, and powder contain from about 5 to 95% protein or other active ingredient of the present invention, and preferably from about 25 to 90% protein or other active ingredient of the present invention. When administered in liquid form, a liquid carrier such as water, petroleum, oils of animal or plant origin such as peanut oil, mineral oil, soybean oil, or sesame oil, or synthetic oils may be added. The liquid form of the pharmaceutical composition may further contain physiological saline solution, dextrose or other saccharide solution, or glycols such as ethylene glycol,

propylene glycol or polyethylene glycol. When administered in liquid form, the pharmaceutical composition contains from about 0.5 to 90% by weight of protein or other active ingredient of the present invention, and preferably from about 1 to 50% protein or other active ingredient of the present invention.

5 When a therapeutically effective amount of protein or other active ingredient of the present invention is administered by intravenous, cutaneous or subcutaneous injection, protein or other active ingredient of the present invention will be in the form of a pyrogen-free, parenterally acceptable aqueous solution. The preparation of such parenterally acceptable protein or other active ingredient solutions, having due regard to  
10 pH, isotonicity, stability, and the like, is within the skill in the art. A preferred pharmaceutical composition for intravenous, cutaneous, or subcutaneous injection should contain, in addition to protein or other active ingredient of the present invention, an isotonic vehicle such as Sodium Chloride Injection, Ringer's Injection, Dextrose Injection, Dextrose and Sodium Chloride Injection, Lactated Ringer's Injection, or other  
15 vehicle as known in the art. The pharmaceutical composition of the present invention may also contain stabilizers, preservatives, buffers, antioxidants, or other additives known to those of skill in the art. For injection, the agents of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiological saline buffer. For transmucosal  
20 administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

For oral administration, the compounds can be formulated readily by combining the active compounds with pharmaceutically acceptable carriers well known in the art. Such carriers enable the compounds of the invention to be formulated as tablets, pills,  
25 dragees, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated. Pharmaceutical preparations for oral use can be obtained from a solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are, in particular, fillers such as sugars, including  
30 lactose, sucrose, mannitol, or sorbitol; cellulose preparations such as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose,

hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired, disintegrating agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate. Dragee cores are provided with suitable coatings. For this purpose, concentrated sugar solutions may be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of active compound doses.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. The push-fit capsules can contain the active ingredients in admixture with filler such as lactose, binders such as starches, and/or lubricants such as talc or magnesium stearate and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycols. In addition, stabilizers may be added. All formulations for oral administration should be in dosages suitable for such administration. For buccal administration, the compositions may take the form of tablets or lozenges formulated in conventional manner.

For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a suitable propellant, *e.g.*, dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, *e.g.*, gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch. The compounds may be formulated for parenteral administration by injection, *e.g.*, by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, *e.g.*, in ampules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or

aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents.

Pharmaceutical formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions.

Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions. Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, *e.g.*, sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, *e.g.*, containing conventional suppository bases such as cocoa butter or other glycerides. In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

A pharmaceutical carrier for the hydrophobic compounds of the invention is a co-solvent system comprising benzyl alcohol, a nonpolar surfactant, a water-miscible organic polymer, and an aqueous phase. The co-solvent system may be the VPD co-solvent system. VPD is a solution of 3% w/v benzyl alcohol, 8% w/v of the nonpolar surfactant polysorbate 80, and 65% w/v polyethylene glycol 300, made up to volume in absolute ethanol. The VPD co-solvent system (VPD:5W) consists of VPD diluted 1:1 with a 5% dextrose in water solution. This co-solvent system dissolves hydrophobic compounds well, and itself produces low toxicity upon systemic administration. Naturally, the proportions of a co-solvent system may be varied considerably without

destroying its solubility and toxicity characteristics. Furthermore, the identity of the co-solvent components may be varied: for example, other low-toxicity nonpolar surfactants may be used instead of polysorbate 80; the fraction size of polyethylene glycol may be varied; other biocompatible polymers may replace polyethylene glycol, e.g. polyvinyl pyrrolidone; and other sugars or polysaccharides may substitute for dextrose. Alternatively, other delivery systems for hydrophobic pharmaceutical compounds may be employed. Liposomes and emulsions are well known examples of delivery vehicles or carriers for hydrophobic drugs. Certain organic solvents such as dimethylsulfoxide also may be employed, although usually at the cost of greater toxicity. Additionally, the compounds may be delivered using a sustained-release system, such as semipermeable matrices of solid hydrophobic polymers containing the therapeutic agent. Various types of sustained-release materials have been established and are well known by those skilled in the art. Sustained-release capsules may, depending on their chemical nature, release the compounds for a few weeks up to over 100 days. Depending on the chemical nature and the biological stability of the therapeutic reagent, additional strategies for protein or other active ingredient stabilization may be employed.

The pharmaceutical compositions also may comprise suitable solid or gel phase carriers or excipients. Examples of such carriers or excipients include but are not limited to calcium carbonate, calcium phosphate, various sugars, starches, cellulose derivatives, gelatin, and polymers such as polyethylene glycols. Many of the active ingredients of the invention may be provided as salts with pharmaceutically compatible counter ions. Such pharmaceutically acceptable base addition salts are those salts which retain the biological effectiveness and properties of the free acids and which are obtained by reaction with inorganic or organic bases such as sodium hydroxide, magnesium hydroxide, ammonia, trialkylamine, dialkylamine, monoalkylamine, dibasic amino acids, sodium acetate, potassium benzoate, triethanol amine and the like.

The pharmaceutical composition of the invention may be in the form of a complex of the protein(s) or other active ingredient(s) of present invention along with protein or peptide antigens. The protein and/or peptide antigen will deliver a stimulatory signal to both B and T lymphocytes. B lymphocytes will respond to antigen through their surface immunoglobulin receptor. T lymphocytes will respond to antigen through the T



cell receptor (TCR) following presentation of the antigen by MHC proteins. MHC and structurally related proteins including those encoded by class I and class II MHC genes on host cells will serve to present the peptide antigen(s) to T lymphocytes. The antigen components could also be supplied as purified MHC-peptide complexes alone or with  
5 co-stimulatory molecules that can directly signal T cells. Alternatively antibodies able to bind surface immunoglobulin and other molecules on B cells as well as antibodies able to bind the TCR and other molecules on T cells can be combined with the pharmaceutical composition of the invention.

The pharmaceutical composition of the invention may be in the form of a  
10 liposome in which protein of the present invention is combined, in addition to other pharmaceutically acceptable carriers, with amphipathic agents such as lipids which exist in aggregated form as micelles, insoluble monolayers, liquid crystals, or lamellar layers in aqueous solution. Suitable lipids for liposomal formulation include, without limitation, monoglycerides, diglycerides, sulfatides, lysolecithins, phospholipids, saponin, bile acids,  
15 and the like. Preparation of such liposomal formulations is within the level of skill in the art, as disclosed, for example, in U.S. Patent Nos. 4,235,871; 4,501,728; 4,837,028; and 4,737,323, all of which are incorporated herein by reference.

The amount of protein or other active ingredient of the present invention in the pharmaceutical composition of the present invention will depend upon the nature and  
20 severity of the condition being treated, and on the nature of prior treatments which the patient has undergone. Ultimately, the attending physician will decide the amount of protein or other active ingredient of the present invention with which to treat each individual patient. Initially, the attending physician will administer low doses of protein or other active ingredient of the present invention and observe the patient's response.

25 Larger doses of protein or other active ingredient of the present invention may be administered until the optimal therapeutic effect is obtained for the patient, and at that point the dosage is not increased further. It is contemplated that the various pharmaceutical compositions used to practice the method of the present invention should contain about 0.01  $\mu$ g to about 100 mg (preferably about 0.1  $\mu$ g to about 10 mg, more  
30 preferably about 0.1  $\mu$ g to about 1 mg) of protein or other active ingredient of the present invention per kg body weight. For compositions of the present invention which are



agent, such as carboxymethyl cellulose or autologous blood clot, to prevent the protein compositions from disassociating from the matrix.

A preferred family of sequestering agents is cellulosic materials such as alkylcelluloses (including hydroxyalkylcelluloses), including methylcellulose, ethylcellulose, hydroxyethylcellulose, hydroxypropylcellulose, hydroxypropyl-methylcellulose, and carboxymethylcellulose, the most preferred being cationic salts of carboxymethylcellulose (CMC). Other preferred sequestering agents include hyaluronic acid, sodium alginate, poly(ethylene glycol), polyoxyethylene oxide, carboxyvinyl polymer and poly(vinyl alcohol). The amount of sequestering agent useful herein is 0.5-20 wt %, preferably 1-10 wt % based on total formulation weight, which represents the amount necessary to prevent desorption of the protein from the polymer matrix and to provide appropriate handling of the composition, yet not so much that the progenitor cells are prevented from infiltrating the matrix, thereby providing the protein the opportunity to assist the osteogenic activity of the progenitor cells. In further compositions, proteins or other active ingredients of the invention may be combined with other agents beneficial to the treatment of the bone and/or cartilage defect, wound, or tissue in question. These agents include various growth factors such as epidermal growth factor (EGF), platelet derived growth factor (PDGF), transforming growth factors (TGF- $\alpha$  and TGF- $\beta$ ), and insulin-like growth factor (IGF).

The therapeutic compositions are also presently valuable for veterinary applications. Particularly domestic animals and thoroughbred horses, in addition to humans, are desired patients for such treatment with proteins or other active ingredients of the present invention. The dosage regimen of a protein-containing pharmaceutical composition to be used in tissue regeneration will be determined by the attending physician considering various factors which modify the action of the proteins, *e.g.*, amount of tissue weight desired to be formed, the site of damage, the condition of the damaged tissue, the size of a wound, type of damaged tissue (*e.g.*, bone), the patient's age, sex, and diet, the severity of any infection, time of administration and other clinical factors. The dosage may vary with the type of matrix used in the reconstitution and with inclusion of other proteins in the pharmaceutical composition. For example, the addition of other known growth factors, such as IGF I (insulin like growth factor I), to the final

composition, may also effect the dosage. Progress can be monitored by periodic assessment of tissue/bone growth and/or repair, for example, X-rays, histomorphometric determinations and tetracycline labeling.

Polynucleotides of the present invention can also be used for gene therapy. Such polynucleotides can be introduced either in vivo or ex vivo into cells for expression in a mammalian subject. Polynucleotides of the invention may also be administered by other known methods for introduction of nucleic acid into a cell or organism (including, without limitation, in the form of viral vectors or naked DNA). Cells may also be cultured ex vivo in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced in vivo for therapeutic purposes.

#### 4.9.3 EFFECTIVE DOSAGE

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve its intended purpose. More specifically, a therapeutically effective amount means an amount effective to prevent development of or to alleviate the existing symptoms of the subject being treated. Determination of the effective amount is well within the capability of those skilled in the art, especially in light of the detailed disclosure provided herein. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from appropriate in vitro assays. For example, a dose can be formulated in animal models to achieve a circulating concentration range that can be used to more accurately determine useful doses in humans. For example, a dose can be formulated in animal models to achieve a circulating concentration range that includes the  $IC_{50}$  as determined in cell culture (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of the protein's biological activity). Such information can be used to more accurately determine useful doses in humans.

A therapeutically effective dose refers to that amount of the compound that results in amelioration of symptoms or a prolongation of survival in a patient. Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical

procedures in cell cultures or experimental animals, *e.g.*, for determining the LD<sub>50</sub> (the dose lethal to 50% of the population) and the ED<sub>50</sub> (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio between LD<sub>50</sub> and ED<sub>50</sub>.

5 Compounds which exhibit high therapeutic indices are preferred. The data obtained from these cell culture assays and animal studies can be used in formulating a range of dosage for use in human. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED<sub>50</sub> with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. The exact formulation, route of administration and dosage can be  
10 chosen by the individual physician in view of the patient's condition. See, *e.g.*, Fingl et al., 1975, in "The Pharmacological Basis of Therapeutics", Ch. 1 p.1. Dosage amount and interval may be adjusted individually to provide plasma levels of the active moiety which are sufficient to maintain the desired effects, or minimal effective concentration  
15 (MEC). The MEC will vary for each compound but can be estimated from *in vitro* data. Dosages necessary to achieve the MEC will depend on individual characteristics and route of administration. However, HPLC assays or bioassays can be used to determine plasma concentrations.

Dosage intervals can also be determined using MEC value. Compounds should  
20 be administered using a regimen which maintains plasma levels above the MEC for 10-90% of the time, preferably between 30-90% and most preferably between 50-90%. In cases of local administration or selective uptake, the effective local concentration of the drug may not be related to plasma concentration.

An exemplary dosage regimen for polypeptides or other compositions of the  
25 invention will be in the range of about 0.01 µg/kg to 100 mg/kg of body weight daily, with the preferred dose being about 0.1 µg/kg to 25 mg/kg of patient body weight daily, varying in adults and children. Dosing may be once daily, or equivalent doses may be delivered at longer or shorter intervals.

The amount of composition administered will, of course, be dependent on the  
30 subject being treated, on the subject's age and weight, the severity of the affliction, the manner of administration and the judgment of the prescribing physician.

#### 4.9.4 PACKAGING

The compositions may, if desired, be presented in a pack or dispenser device which may contain one or more unit dosage forms containing the active ingredient. The pack may, for example, comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration. Compositions comprising a compound of the invention formulated in a compatible pharmaceutical carrier may also be prepared, placed in an appropriate container, and labeled for treatment of an indicated condition.

#### 4.10 ANTIBODIES

Another aspect of the invention is an antibody that specifically binds the polypeptide of the invention. Such antibodies include monoclonal and polyclonal antibodies, single chain antibodies, chimeric antibodies, bifunctional/bispecific antibodies, humanized antibodies, human antibodies, and complementary determining region (CDR)-grafted antibodies, including compounds which include CDR and/or antigen-binding sequences, which specifically recognize a polypeptide of the invention. Preferred antibodies of the invention are human antibodies which are produced and identified according to methods described in WO93/11236, published June 20, 1993, which is incorporated herein by reference in its entirety. Antibody fragments, including Fab, Fab', F(ab')<sub>2</sub>, and F<sub>v</sub>, are also provided by the invention. The term "specific for" indicates that the variable regions of the antibodies of the invention recognize and bind polypeptides of the invention exclusively (*i.e.*, able to distinguish the polypeptide of the invention from other similar polypeptides despite sequence identity, homology, or similarity found in the family of polypeptides), but may also interact with other proteins (for example, *S. aureus* protein A or other antibodies in ELISA techniques) through interactions with sequences outside the variable region of the antibodies, and in particular, in the constant region of the molecule. Screening assays to determine binding specificity of an antibody of the invention are well known and routinely practiced in the art. For a comprehensive discussion of such assays, see Harlow et al. (Eds), Antibodies A Laboratory Manual; Cold Spring Harbor Laboratory; Cold Spring Harbor, NY (1988),

Chapter 6. Antibodies that recognize and bind fragments of the polypeptides of the invention are also contemplated, provided that the antibodies are first and foremost specific for, as defined above, full length polypeptides of the invention. As with antibodies that are specific for full length polypeptides of the invention, antibodies of the invention that recognize fragments are those which can distinguish polypeptides from the same family of polypeptides despite inherent sequence identity, homology, or similarity found in the family of proteins. Antibodies of the invention can be produced using any method well known and routinely practiced in the art.

Non-human antibodies may be humanized by any methods known in the art. In one method, the non-human CDRs are inserted into a human antibody or consensus antibody framework sequence. Further changes can then be introduced into the antibody framework to modulate affinity or immunogenicity.

Antibodies of the invention are useful for, for example, therapeutic purposes (by modulating activity of a polypeptide of the invention), diagnostic purposes to detect or quantitate a polypeptide of the invention, as well as purification of a polypeptide of the invention. Kits comprising an antibody of the invention for any of the purposes described herein are also comprehended. In general, a kit of the invention also includes a control antigen for which the antibody is immunospecific. The invention further provides a hybridoma that produces an antibody according to the invention. Antibodies of the invention are useful for detection and/or purification of the polypeptides of the invention.

Polypeptides of the invention may also be used to immunize animals to obtain polyclonal and monoclonal antibodies which specifically react with the protein. Such antibodies may be obtained using either the entire protein or fragments thereof as an immunogen. The peptide immunogens additionally may contain a cysteine residue at the carboxyl terminus, and are conjugated to a hapten such as keyhole limpet hemocyanin (KLH). Methods for synthesizing such peptides are known in the art, for example, as in R. P. Merrifield, J. Amer. Chem. Soc. 85, 2149-2154 (1963); J. L. Krstenansky, et al., FEBS Lett. 211, 10 (1987).

Monoclonal antibodies binding to the protein of the invention may be useful diagnostic agents for the immunodetection of the protein. Neutralizing monoclonal antibodies binding to the protein may also be useful therapeutics for both conditions

associated with the protein and also in the treatment of some forms of cancer where abnormal expression of the protein is involved. In the case of cancerous cells or leukemic cells, neutralizing monoclonal antibodies against the protein may be useful in detecting and preventing the metastatic spread of the cancerous cells, which may be mediated by the protein. In general, techniques for preparing polyclonal and monoclonal antibodies as well as hybridomas capable of producing the desired antibody are well known in the art (Campbell, A.M., *Monoclonal Antibodies Technology: Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1984); St. Groth et al., *J. Immunol.* 35:1-21 (1990); Kohler and Milstein, *Nature* 256:495-497 (1975)), the trioma technique, the human B-cell hybridoma technique (Kozbor et al., *Immunology Today* 4:72 (1983); Cole et al., in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc. (1985), pp. 77-96).

Any animal (mouse, rabbit, *etc.*) which is known to produce antibodies can be immunized with a peptide or polypeptide of the invention. Methods for immunization are well known in the art. Such methods include subcutaneous or intraperitoneal injection of the polypeptide. One skilled in the art will recognize that the amount of the protein encoded by the ORF of the present invention used for immunization will vary based on the animal which is immunized, the antigenicity of the peptide and the site of injection. The protein that is used as an immunogen may be modified or administered in an adjuvant in order to increase the protein's antigenicity. Methods of increasing the antigenicity of a protein are well known in the art and include, but are not limited to, coupling the antigen with a heterologous protein (such as globulin or  $\beta$ -galactosidase) or through the inclusion of an adjuvant during immunization.

For monoclonal antibodies, spleen cells from the immunized animals are removed, fused with myeloma cells, such as SP2/0-Ag14 myeloma cells, and allowed to become monoclonal antibody producing hybridoma cells. Any one of a number of methods well known in the art can be used to identify the hybridoma cell which produces an antibody with the desired characteristics. These include screening the hybridomas with an ELISA assay, Western blot analysis, or radioimmunoassay (Lutz et al., *Exp. Cell Research*. 175:109-124 (1988)). Hybridomas secreting the desired antibodies are cloned and the class and subclass is determined using procedures known in the art (Campbell,



A.M., Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1984)). Techniques described for the production of single chain antibodies (U.S. Patent 4,946,778) can be adapted to produce single chain antibodies to proteins of the present invention.

For polyclonal antibodies, antibody-containing antiserum is isolated from the immunized animal and is screened for the presence of antibodies with the desired specificity using one of the above-described procedures. The present invention further provides the above-described antibodies in delectably labeled form. Antibodies can be delectably labeled through the use of radioisotopes, affinity labels (such as biotin, avidin, etc.), enzymatic labels (such as horseradish peroxidase, alkaline phosphatase, etc.) fluorescent labels (such as FITC or rhodamine, etc.), paramagnetic atoms, etc. Procedures for accomplishing such labeling are well-known in the art, for example, see (Sternberger, L.A. et al., J. Histochem. Cytochem. 18:315 (1970); Bayer, E.A. et al., Meth. Enzym. 62:308 (1979); Engval, E. et al., Immunol. 109:129 (1972); Goding, J.W. J. Immunol. Meth. 13:215 (1976)).

The labeled antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays to identify cells or tissues in which a fragment of the polypeptide of interest is expressed. The antibodies may also be used directly in therapies or other diagnostics. The present invention further provides the above-described antibodies immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such as agarose and Sepharose®, acrylic resins and such as polyacrylamide and latex beads. Techniques for coupling antibodies to such solid supports are well known in the art (Weir, D.M. et al., "Handbook of Experimental Immunology" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10 (1986); Jacoby, W.D. et al., Meth. Enzym. 34 Academic Press, N.Y. (1974)). The immobilized antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays as well as for immuno-affinity purification of the proteins of the present invention.

#### 4.11 COMPUTER READABLE SEQUENCES

In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium which can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide sequence of the present invention. As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate manufactures comprising the nucleotide sequence information of the present invention.

A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data processor structuring formats (*e.g.* text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

By providing any of the nucleotide sequences SEQ ID NOs: 1 - 35 or a representative fragment thereof; or a nucleotide sequence at least 95% identical to any of the nucleotide sequences of the SEQ ID NOs: 1 - 35 in computer readable form, a skilled artisan can routinely access the sequence information for a variety of purposes. Computer software is publicly available which allows a skilled artisan to access sequence

information provided in a computer readable medium. The examples which follow demonstrate how software which implements the BLAST (Altschul et al., J. Mol. Biol. 215:403-410 (1990)) and BLAZE (Brutlag et al., Comp. Chem. 17:203-207 (1993)) search algorithms on a Sybase system is used to identify open reading frames (ORFs)

5 within a nucleic acid sequence. Such ORFs may be protein encoding fragments and may be useful in producing commercially important proteins such as enzymes used in fermentation reactions and in the production of commercially useful metabolites.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence  
10 information of the present invention. The minimum hardware means of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based systems are suitable for use in the present invention. As stated above, the computer-based systems of the present  
15 invention comprise a data storage means having stored therein a nucleotide sequence of the present invention and the necessary hardware means and software means for supporting and implementing a search means. As used herein, "data storage means" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means which can access manufactures having recorded  
20 thereon the nucleotide sequence information of the present invention.

As used herein, "search means" refers to one or more programs which are implemented on the computer-based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of a known sequence which  
25 match a particular target sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, Smith-Waterman, MacPattern (EMBL), BLASTN and BLASTA (NPOLYPEPTIDEIA). A skilled artisan can readily  
30 recognize that any one of the available algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based

systems. As used herein, a "target sequence" can be any nucleic acid or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 300 amino acids, more preferably from about 30 to 100 nucleotide residues. However, it is well recognized that searches for commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzyme active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

#### 4.12 TRIPLE HELIX FORMATION

In addition, the fragments of the present invention, as broadly described, can be used to control gene expression through triple helix formation or antisense DNA or RNA, both of which methods are based on the binding of a polynucleotide sequence to DNA or RNA. Polynucleotides suitable for use in these methods are preferably 20 to 40 bases in length and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 15241:456 (1988); and Dervan et al., Science 251:1360 (1991)) or to the mRNA itself (antisense - Olmno, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide.

#### 4.13 DIAGNOSTIC ASSAYS AND KITS

The present invention further provides methods to identify the presence or expression of one of the ORFs of the present invention, or homolog thereof, in a test sample, using a nucleic acid probe or antibodies of the present invention, optionally conjugated or otherwise associated with a suitable label.

In general, methods for detecting a polynucleotide of the invention can comprise contacting a sample with a compound that binds to and forms a complex with the polynucleotide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polynucleotide of the invention is detected in the sample. Such methods can also comprise contacting a sample under stringent hybridization conditions with nucleic acid primers that anneal to a polynucleotide of the invention under such conditions, and amplifying annealed polynucleotides, so that if a polynucleotide is amplified, a polynucleotide of the invention is detected in the sample.

In general, methods for detecting a polypeptide of the invention can comprise contacting a sample with a compound that binds to and forms a complex with the polypeptide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polypeptide of the invention is detected in the sample.

In detail, such methods comprise incubating a test sample with one or more of the antibodies or one or more of the nucleic acid probes of the present invention and assaying for binding of the nucleic acid probes or antibodies to components within the test sample.

Conditions for incubating a nucleic acid probe or antibody with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the nucleic acid probe or antibody used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or immunological assay formats can readily be adapted to employ the nucleic acid probes or antibodies of the present invention. Examples of such assays can be found in Chard, T., *An Introduction to Radioimmunoassay and Related Techniques*, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G.R. et al., *Techniques in Immunocytochemistry*, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., *Practice and Theory of immunoassays*:

Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1985). The test samples of the present invention include cells, protein or membrane extracts of cells, or biological fluids such as sputum, blood, serum, plasma, or urine. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can be readily be adapted in order to obtain a sample which is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention. Specifically, the invention provides a compartment kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the probes or antibodies of the present invention; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound probe or antibody.

In detail, a compartment kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the antibodies used in the assay, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound antibody or probe. Types of detection reagents include labeled nucleic acid probes, labeled secondary antibodies, or in the alternative, if the primary antibody is labeled, the enzymatic, or antibody binding reagents which are capable of reacting with the labeled antibody. One skilled in the art will readily recognize that the disclosed probes and antibodies of the present invention can be readily incorporated into one of the established kit formats which are well known in the art.

#### 4.14 MEDICAL IMAGING

The novel polypeptides and binding partners of the invention are useful in medical imaging of sites expressing the molecules of the invention (e.g., where the polypeptide of the invention is involved in the immune response, for imaging sites of inflammation or infection). See, e.g., Kunkel et al., U.S. Pat. NO. 5,413,778. Such methods involve chemical attachment of a labeling or imaging agent, administration of the labeled polypeptide to a subject in a pharmaceutically acceptable carrier, and imaging the labeled polypeptide *in vivo* at the target site.

#### 4.15 SCREENING ASSAYS

Using the isolated proteins and polynucleotides of the invention, the present invention further provides methods of obtaining and identifying agents which bind to a polypeptide encoded by an ORF corresponding to any of the nucleotide sequences set forth in the SEQ ID NOs: 1 - 35, or bind to a specific domain of the polypeptide encoded by the nucleic acid. In detail, said method comprises the steps of:

- (a) contacting an agent with an isolated protein encoded by an ORF of the present invention, or nucleic acid of the invention; and
- (b) determining whether the agent binds to said protein or said nucleic acid.

In general, therefore, such methods for identifying compounds that bind to a polynucleotide of the invention can comprise contacting a compound with a polynucleotide of the invention for a time sufficient to form a polynucleotide/compound complex, and detecting the complex, so that if a polynucleotide/compound complex is detected, a compound that binds to a polynucleotide of the invention is identified.

Likewise, in general, therefore, such methods for identifying compounds that bind to a polypeptide of the invention can comprise contacting a compound with a polypeptide of the invention for a time sufficient to form a polypeptide/compound complex, and detecting the complex, so that if a polypeptide/compound complex is detected, a compound that binds to a polynucleotide of the invention is identified.

Methods for identifying compounds that bind to a polypeptide of the invention can also comprise contacting a compound with a polypeptide of the invention in a cell for a time sufficient to form a polypeptide/compound complex, wherein the complex drives

expression of a receptor gene sequence in the cell, and detecting the complex by detecting reporter gene sequence expression, so that if a polypeptide/compound complex is detected, a compound that binds a polypeptide of the invention is identified.

Compounds identified via such methods can include compounds which modulate the activity of a polypeptide of the invention (that is, increase or decrease its activity, relative to activity observed in the absence of the compound). Alternatively, compounds identified via such methods can include compounds which modulate the expression of a polynucleotide of the invention (that is, increase or decrease expression relative to expression levels observed in the absence of the compound). Compounds, such as compounds identified via the methods of the invention, can be tested using standard assays well known to those of skill in the art for their ability to modulate activity/expression.

The agents screened in the above assay can be, but are not limited to, peptides, carbohydrates, vitamin derivatives, or other pharmaceutical agents. The agents can be selected and screened at random or rationally selected or designed using protein modeling techniques.

For random screening, agents such as peptides, carbohydrates, pharmaceutical agents and the like are selected at random and are assayed for their ability to bind to the protein encoded by the ORF of the present invention. Alternatively, agents may be rationally selected or designed. As used herein, an agent is said to be "rationally selected or designed" when the agent is chosen based on the configuration of the particular protein. For example, one skilled in the art can readily adapt currently available procedures to generate peptides, pharmaceutical agents and the like, capable of binding to a specific peptide sequence, in order to generate rationally designed antipeptide peptides, for example see Hurby et al., Application of Synthetic Peptides: Antisense Peptides," In Synthetic Peptides, A User's Guide, W.H. Freeman, NY (1992), pp. 289-307, and Kaspczak et al., Biochemistry 28:9230-8 (1989), or pharmaceutical agents, or the like.

In addition to the foregoing, one class of agents of the present invention, as broadly described, can be used to control gene expression through binding to one of the ORFs or EMFs of the present invention. As described above, such agents can be randomly screened or rationally designed/selected. Targeting the ORF or EMF allows a



skilled artisan to design sequence specific or element specific agents, modulating the expression of either a single ORF or multiple ORFs which rely on the same EMF for expression control. One class of DNA binding agents are agents which contain base residues which hybridize or form a triple helix formation by binding to DNA or RNA.

- 5 Such agents can be based on the classic phosphodiester, ribonucleic acid backbone, or can be a variety of sulfhydryl or polymeric derivatives which have base attachment capacity.

Agents suitable for use in these methods preferably contain 20 to 40 bases and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 241:456 (1988); and Dervan et al., Science 251:1360 (1991)) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide and other DNA binding agents.

Agents which bind to a protein encoded by one of the ORFs of the present invention can be used as a diagnostic agent. Agents which bind to a protein encoded by one of the ORFs of the present invention can be formulated using known techniques to generate a pharmaceutical composition.

#### 4.16 USE OF NUCLEIC ACIDS AS PROBES

Another aspect of the subject invention is to provide for polypeptide-specific nucleic acid hybridization probes capable of hybridizing with naturally occurring nucleotide sequences. The hybridization probes of the subject invention may be derived from any of the nucleotide sequences SEQ ID NOs: 1 - 35. Because the corresponding gene is only expressed in a limited number of tissues, a hybridization probe derived from any of the nucleotide sequences SEQ ID NOs: 1 - 35 can be used as an indicator of the presence of RNA of cell type of such a tissue in a sample.

Any suitable hybridization technique can be employed, such as, for example, in situ hybridization. PCR as described in US Patents Nos. 4,683,195 and 4,965,188 provides additional uses for oligonucleotides based upon the nucleotide sequences. Such probes used in PCR may be of recombinant origin, may be chemically synthesized, or a mixture of both. The probe will comprise a discrete nucleotide sequence for the detection of identical sequences or a degenerate pool of possible sequences for identification of closely related genomic sequences.

Other means for producing specific hybridization probes for nucleic acids include the cloning of nucleic acid sequences into vectors for the production of mRNA probes.

Such vectors are known in the art and are commercially available and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerase as T7 or SP6 RNA polymerase and the appropriate radioactively labeled nucleotides. The nucleotide sequences may be used to construct hybridization probes for mapping their respective genomic sequences. The nucleotide sequence provided herein may be mapped to a chromosome or specific regions of a chromosome using well known genetic and/or chromosomal mapping techniques. These techniques include in situ hybridization, linkage analysis against known chromosomal markers, hybridization screening with libraries or flow-sorted chromosomal preparations specific to known chromosomes, and the like. The technique of fluorescent in situ hybridization of chromosome spreads has been described, among other places, in Verma et al (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York NY.

Fluorescent *in situ* hybridization of chromosomal preparations and other physical chromosome mapping techniques may be correlated with additional genetic map data. Examples of genetic map data can be found in the 1994 Genome Issue of Science (265:1981f). Correlation between the location of a nucleic acid on a physical chromosomal map and a specific disease (or predisposition to a specific disease) may help delimit the region of DNA associated with that genetic disease. The nucleotide sequences of the subject invention may be used to detect differences in gene sequences between normal, carrier or affected individuals.

#### 4.17 PREPARATION OF SUPPORT BOUND OLIGONUCLEOTIDES

Oligonucleotides, i.e., small nucleic acid segments, may be readily prepared by, for example, directly synthesizing the oligonucleotide by chemical means, as is commonly practiced using an automated oligonucleotide synthesizer.

Support bound oligonucleotides may be prepared by any of the methods known to those of skill in the art using any suitable support such as glass, polystyrene or Teflon. One strategy is to precisely spot oligonucleotides synthesized by standard synthesizers.

Immobilization can be achieved using passive adsorption (Inouye & Hondo, (1990) J. Clin. Microbiol. 28(6) 1469-72); using UV light (Nagata *et al.*, 1985; Dahlen *et al.*, 1987; Morrissey & Collins, (1989) Mol. Cell Probes 3(2) 189-207) or by covalent binding of base modified DNA (Keller *et al.*, 1988; 1989); all references being specifically incorporated herein.

Another strategy that may be employed is the use of the strong biotin-streptavidin interaction as a linker. For example, Broude *et al.* (1994) Proc. Natl. Acad. Sci. USA 91(8) 3072-6, describe the use of biotinylated probes, although these are duplex probes, that are immobilized on streptavidin-coated magnetic beads. Streptavidin-coated beads may be purchased from Dynal, Oslo. Of course, this same linking chemistry is applicable to coating any surface with streptavidin. Biotinylated probes may be purchased from various sources, such as, e.g., Operon Technologies (Alameda, CA).

Nunc Laboratories (Naperville, IL) is also selling suitable material that could be used. Nunc Laboratories have developed a method by which DNA can be covalently bound to the microwell surface termed CovaLink NH. CovaLink NH is a polystyrene surface grafted with secondary amino groups (>NH) that serve as bridge-heads for further covalent coupling. CovaLink Modules may be purchased from Nunc Laboratories. DNA molecules may be bound to CovaLink exclusively at the 5'-end by a phosphoramidate bond, allowing immobilization of more than 1 pmol of DNA (Rasmussen *et al.*, (1991) Anal. Biochem. 198(1) 138-42).

The use of CovaLink NH strips for covalent binding of DNA molecules at the 5'-end has been described (Rasmussen *et al.*, (1991). In this technology, a phosphoramidate bond is employed (Chu *et al.*, (1983) Nucleic Acids Res. 11(8) 6513-29). This is beneficial as immobilization using only a single covalent bond is preferred. The phosphoramidate bond

joins the DNA to the CovaLink NH secondary amino groups that are positioned at the end of spacer arms covalently grafted onto the polystyrene surface through a 2 nm long spacer arm. To link an oligonucleotide to CovaLink NH via an phosphoramidate bond, the oligonucleotide terminus must have a 5'-end phosphate group. It is, perhaps, even possible for biotin to be covalently bound to CovaLink and then streptavidin used to bind the probes.

More specifically, the linkage method includes dissolving DNA in water (7.5 ng/ul) and denaturing for 10 min. at 95°C and cooling on ice for 10 min. Ice-cold 0.1 M 1-methylimidazole, pH 7.0 (1-MeIm<sub>7</sub>), is then added to a final concentration of 10 mM 1-MeIm<sub>7</sub>. A ss DNA solution is then dispensed into CovaLink NH strips (75 ul/well) standing on ice.

Carbodiimide 0.2 M 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC), dissolved in 10 mM 1-MeIm<sub>7</sub>, is made fresh and 25 ul added per well. The strips are incubated for 5 hours at 50°C. After incubation the strips are washed using, e.g., Nunc-Immuno Wash; first the wells are washed 3 times, then they are soaked with washing solution for 5 min., and finally they are washed 3 times (where in the washing solution is 0.4 N NaOH, 0.25% SDS heated to 50°C).

It is contemplated that a further suitable method for use with the present invention is that described in PCT Patent Application WO 90/03382 (Southern & Maskos), incorporated herein by reference. This method of preparing an oligonucleotide bound to a support involves attaching a nucleoside 3'-reagent through the phosphate group by a covalent phosphodiester link to aliphatic hydroxyl groups carried by the support. The oligonucleotide is then synthesized on the supported nucleoside and protecting groups removed from the synthetic oligonucleotide chain under standard conditions that do not cleave the oligonucleotide from the support. Suitable reagents include nucleoside phosphoramidite and nucleoside hydrogen phosphorate.

An on-chip strategy for the preparation of DNA probe for the preparation of DNA probe arrays may be employed. For example, addressable laser-activated photodeprotection may be employed in the chemical synthesis of oligonucleotides directly on a glass surface, as described by Fodor *et al.* (1991) Science 251(4995) 767-73, incorporated herein by reference. Probes may also be immobilized on nylon supports as described by Van Ness *et al.* (1991) Nucleic Acids Res. 19(12) 3345-50; or linked to Teflon using the method of

Duncan & Cavalier (1988) Anal. Biochem. 169(1) 104-8; all references being specifically incorporated herein.

To link an oligonucleotide to a nylon support, as described by Van Ness *et al.* (1991), requires activation of the nylon surface via alkylation and selective activation of the 5'-amine of oligonucleotides with cyanuric chloride.

One particular way to prepare support bound oligonucleotides is to utilize the light-generated synthesis described by Pease *et al.*, (1994) PNAS USA 91(11) 5022-6, incorporated herein by reference). These authors used current photolithographic techniques to generate arrays of immobilized oligonucleotide probes (DNA chips). These methods, in which light is used to direct the synthesis of oligonucleotide probes in high-density, miniaturized arrays, utilize photolabile 5'-protected *N*-acyl-deoxynucleoside phosphoramidites, surface linker chemistry and versatile combinatorial synthesis strategies. A matrix of 256 spatially defined oligonucleotide probes may be generated in this manner.

#### 4.18 PREPARATION OF NUCLEIC ACID FRAGMENTS

The nucleic acids may be obtained from any appropriate source, such as cDNAs, genomic DNA, chromosomal DNA, microdissected chromosome bands, cosmid or YAC inserts, and RNA, including mRNA without any amplification steps. For example, Sambrook *et al.* (1989) describes three protocols for the isolation of high molecular weight DNA from mammalian cells (p. 9.14-9.23).

DNA fragments may be prepared as clones in M13, plasmid or lambda vectors and/or prepared directly from genomic DNA or cDNA by PCR or other amplification methods. Samples may be prepared or dispensed in multiwell plates. About 100-1000 ng of DNA samples may be prepared in 2-500 ml of final volume.

The nucleic acids would then be fragmented by any of the methods known to those of skill in the art including, for example, using restriction enzymes as described at 9.24-9.28 of Sambrook *et al.* (1989), shearing by ultrasound and NaOH treatment.

Low pressure shearing is also appropriate, as described by Schriefer *et al.* (1990) Nucleic Acids Res. 18(24) 7455-6, incorporated herein by reference). In this method, DNA samples are passed through a small French pressure cell at a variety of low to intermediate pressures. A lever device allows controlled application of low to intermediate pressures to

the cell. The results of these studies indicate that low-pressure shearing is a useful alternative to sonic and enzymatic DNA fragmentation methods.

One particularly suitable way for fragmenting DNA is contemplated to be that using the two base recognition endonuclease, *Cvi*II, described by Fitzgerald *et al.* (1992) Nucleic Acids Res. 20(14) 3753-62. These authors described an approach for the rapid fragmentation and fractionation of DNA into particular sizes that they contemplated to be suitable for shotgun cloning and sequencing.

The restriction endonuclease *Cvi*II normally cleaves the recognition sequence PuGCPy between the G and C to leave blunt ends. Atypical reaction conditions, which alter the specificity of this enzyme (*Cvi*II\*\*), yield a quasi-random distribution of DNA fragments from the small molecule pUC19 (2688 base pairs). Fitzgerald *et al.* (1992) quantitatively evaluated the randomness of this fragmentation strategy, using a *Cvi*II\*\* digest of pUC19 that was size fractionated by a rapid gel filtration method and directly ligated, without end repair, to a lac Z minus M13 cloning vector. Sequence analysis of 76 clones showed that *Cvi*II\*\* restricts pyGCPy and PuGCPu, in addition to PuGCPy sites, and that new sequence data is accumulated at a rate consistent with random fragmentation.

As reported in the literature, advantages of this approach compared to sonication and agarose gel fractionation include: smaller amounts of DNA are required (0.2-0.5 ug instead of 2-5 ug); and fewer steps are involved (no preligation, end repair, chemical extraction, or agarose gel electrophoresis and elution are needed).

Irrespective of the manner in which the nucleic acid fragments are obtained or prepared, it is important to denature the DNA to give single stranded pieces available for hybridization. This is achieved by incubating the DNA solution for 2-5 minutes at 80-90°C. The solution is then cooled quickly to 2°C to prevent renaturation of the DNA fragments before they are contacted with the chip. Phosphate groups must also be removed from genomic DNA by methods known in the art.

#### 4.19 PREPARATION OF DNA ARRAYS

Arrays may be prepared by spotting DNA samples on a support such as a nylon membrane. Spotting may be performed by using arrays of metal pins (the positions of which correspond to an array of wells in a microtiter plate) to repeated by transfer of about 20 nl of a DNA solution to a nylon membrane. By offset printing, a density of dots higher than the

density of the wells is achieved. One to 25 dots may be accommodated in 1 mm<sup>2</sup>, depending on the type of label used. By avoiding spotting in some preselected number of rows and columns, separate subsets (subarrays) may be formed. Samples in one subarray may be the same genomic segment of DNA (or the same gene) from different individuals, or may be different, overlapped genomic clones. Each of the subarrays may represent replica spotting of the same samples. In one example, a selected gene segment may be amplified from 64 patients. For each patient, the amplified gene segment may be in one 96-well plate (all 96 wells containing the same sample). A plate for each of the 64 patients is prepared. By using a 96-pin device, all samples may be spotted on one 8 x 12 cm membrane. Subarrays may contain 64 samples, one from each patient. Where the 96 subarrays are identical, the dot span may be 1 mm<sup>2</sup> and there may be a 1 mm space between subarrays.

Another approach is to use membranes or plates (available from NUNC, Naperville, Illinois) which may be partitioned by physical spacers e.g. a plastic grid molded over the membrane, the grid being similar to the sort of membrane applied to the bottom of multiwell plates, or hydrophobic strips. A fixed physical spacer is not preferred for imaging by exposure to flat phosphor-storage screens or x-ray films.

The present invention is illustrated in the following examples. Upon consideration of the present disclosure, one of skill in the art will appreciate that many other embodiments and variations may be made in the scope of the present invention. Accordingly, it is intended that the broader aspects of the present invention not be limited to the disclosure of the following examples. The present invention is not to be limited in scope by the exemplified embodiments which are intended as illustrations of single aspects of the invention, and compositions and methods which are functionally equivalent are within the scope of the invention. Indeed, numerous modifications and variations in the practice of the invention are expected to occur to those skilled in the art upon consideration of the present preferred embodiments. Consequently, the only limitations which should be placed upon the scope of the invention are those which appear in the appended claims.

All references cited within the body of the instant specification are hereby incorporated by reference in their entirety.

## 5.0 EXAMPLES

### 5.1 EXAMPLE 1

#### Novel Nucleic Acid Sequences Obtained From Various Libraries

A plurality of novel nucleic acids were obtained from cDNA libraries prepared from various human tissues and in some cases isolated from a genomic library derived from human chromosome using standard PCR, SBH sequence signature analysis and Sanger sequencing techniques. The inserts of the library were amplified with PCR using primers specific for the vector sequences which flank the inserts. Clones from cDNA libraries were spotted on nylon membrane filters and screened with oligonucleotide probes (e.g., 7-mers) to obtain signature sequences. The clones were clustered into groups of similar or identical sequences. Representative clones were selected for sequencing.

In some cases, the 5' sequence of the amplified inserts was then deduced using a typical Sanger sequencing protocol. PCR products were purified and subjected to fluorescent dye terminator cycle sequencing. Single pass gel sequencing was done using a 377 Applied Biosystems (ABI) sequencer to obtain the novel nucleic acid sequences. In some cases RACE (Random Amplification of cDNA Ends) was performed to further extend the sequence in the 5' direction.

### 5.2 EXAMPLE 2

#### Novel Nucleic Acids

The novel nucleic acids of the present invention of the invention were assembled from sequences that were obtained from a cDNA library by methods described in Example 1 above, and in some cases sequences obtained from one or more public databases. The nucleic acids were assembled using an EST sequence as a seed. Then a recursive algorithm was used to extend the seed EST into an extended assemblage, by pulling additional sequences from different databases (i.e., Hyseq's database containing EST sequences, dbEST version 114, gb pri 114, and UniGene version 101) that belong to this assemblage. The algorithm terminated when there was no additional sequences from the above databases that would extend the assemblage. Inclusion of component sequences into the assemblage was based on a BLASTN hit to the extending assemblage with BLAST score greater than 300 and percent identity greater than 95%.



Using PHRAP (Univ. of Washington) or CAP4 (Paracel), a full length gene cDNA sequence and its corresponding protein sequence were generated from the assemblage. Any frame shifts and incorrect stop codons were corrected by hand editing. During editing, the sequence was checked using FASTY and/or BLAST against Genbank (i.e., dbEST version 118, gb pri 118, UniGene version 118, Genepet release 118). Other computer programs which may have been used in the editing process were phredPhrap and Consed (University of Washington) and ed-ready, ed-ext and gc-zip-2 (Hyseq, Inc.). The full-length nucleotide and amino acid sequences, including splice variants resulting from these procedures are shown in the Sequence Listing as SEQ ID NOS: 1- 35.

Table 1 shows the various tissue sources of SEQ ID NO: 1-35.

The homology for SEQ ID NO: 1-35 were obtained by a BLASTP version 2.0a1 19MP-WashU search against Genpept release 118, using BLAST algorithm. The results showed homologues for SEQ ID NO: 1-35 from Genpept. The homologues with identifiable functions for SEQ ID NO: 1-35 are shown in Table 2 below.

Using eMatrix software package (Stanford University, Stanford, CA) (Wu et al., J. Comp. Biol., Vol. 6 pp. 219-235 (1999) herein incorporated by reference), all the sequences were examined to determine whether they had identifiable signature regions. Table 3 shows the signature region found in the indicated polypeptide sequences, the description of the signature, the eMatrix p-value(s) and the position(s) of the signature within the polypeptide sequence.

Using the pFam software program (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1) pp. 320-322 (1998) herein incorporated by reference) all the polypeptide sequences were examined for domains with homology to certain peptide domains. Table 4 shows the name of the domain found, the description, the p-value and the pFam score for the identified domain within the sequence.

The nucleotide sequence within the sequences that codes for signal peptide sequences and their cleavage sites can be determine from using Neural Network SignalP V1.1 program (from Center for Biological Sequence Analysis, The Technical University of Denmark). The process for identifying prokaryotic and eukaryotic signal peptides and their cleavage sites are also disclosed by Henrik Nielson, Jacob Engelbrecht, Soren Brunak, and Gunnar von Heijne in the publication " Identification of prokaryotic and

eukaryotic signal peptides and prediction of their cleavage sites” Protein Engineering, Vol. 10, no. 1, pp. 1-6 (1997), incorporated herein by reference. A maximum S score and a mean S score, as described in the Nielson et al as reference, was obtained for the polypeptide sequences. Table 5 shows the position of the signal peptide in each of the polypeptides and the maximum score and mean score associated with that signal peptide.

TABLE 1

TISSUE ORIGIN	RNA SOURCE	HYSEQ LIBRARY NAME	SEQ ID NOS:
adult brain	GIBCO	AB3001	27
adult brain	GIBCO	ABD003	2 5 23-24 30-32
adult brain	Clontech	ABR006	2 12 31
adult brain	Clontech	ABR008	5-8 10-11 13 20 23 27 30-31 34-35
adult brain	Invitrogen	ABR014	24
adult brain	Invitrogen	ABT004	11 30-31
cultured preadipocytes	Strategene	ADP001	7-8 16-17 30
adrenal gland	Clontech	ADR002	29 31
adult heart	GIBCO	AHR001	3-4 7-8 13 16-17 23 26 31-32 34-35
adult kidney	GIBCO	AKD001	1-2 4 7-8 10 18 23 30-31 35
adult kidney	Invitrogen	AKT002	1 15-17 23
adult lung	GIBCO	ALG001	23 26
lymph node	Clontech	ALN001	6 25
young liver	GIBCO	ALV001	12 19 27 32 35
adult liver	Invitrogen	ALV002	4 12 23 25 27 31
adult liver	Clontech	ALV003	12
adult ovary	Invitrogen	AOV001	1 4 7-8 10 13 15-19 23-24 28 30-32 35
adult placenta	Clontech	APL001	11
placenta	Invitrogen	APL002	6 28 30
adult spleen	GIBCO	ASP001	6 16-17 31 35
testis	GIBCO	ATS001	2 15-17 19 23 31-32 35
adult bladder	Invitrogen	BLD001	27
bone marrow	Clontech	BMD001	2 4 6 11 16-17 22 24-25 29 32-35
bone marrow	Clontech	BMD002	2 4-6 13 16-17 25 31 35
bone marrow	Clontech	BMD007	1
adult colon	Invitrogen	CLN001	5 34
Mixture of 16 tissues - mRNAs*	Various Vendors*	CTL016	33
adult cervix	BioChain	CVX001	4 7-8 15-17 21 25 28
diaphragm	BioChain	DIA002	16-17
endothelial cells	Strategene	EDT001	4-5 7-8 14 16-17 19 23 28 30-31 33-35
fetal brain	Clontech	FBR006	6-8 10-11 20 23 29 32 34-35
fetal brain	Invitrogen	FBT002	5-8 10 23
fetal heart	Invitrogen	FHR001	23
fetal kidney	Clontech	FKD001	10

\* The 16 tissue-mRNAs and their vendor source, are as follows: 1) Normal adult brain mRNA (Invitrogen), 2) normal adult kidney mRNA (Invitrogen), 3) normal adult liver mRNA (Invitrogen), 4) normal fetal brain mRNA (Invitrogen), 5) normal fetal kidney mRNA (Invitrogen), 6) normal fetal liver mRNA (Invitrogen), 7) normal fetal skin mRNA (Invitrogen), 8) human adrenal gland mRNA (Clontech), 9) human bone marrow mRNA (Clontech), 10) human leukemia lymphoblastic mRNA (Clontech), 11) human thymus mRNA (Clontech), 12) human lymph node mRNA (Clontech), 13) human spinal cord mRNA (Clontech), 14) human thyroid mRNA (Clontech), 15) human esophagus mRNA (BioChain), 16) human conceptional umbilical cord mRNA (BioChain).

TABLE 1

TISSUE ORIGIN	RNA SOURCE	HYSEQ LIBRARY NAME	SEQ ID NOS:
fetal lung	Clontech	FLG001	12 31
fetal lung	Invitrogen	FLG003	16-17 31
fetal liver-spleen	Columbia University	FLS001	1 4 6-8 10 12 18-19 23-25 28 30-35
fetal liver-spleen	Columbia University	FLS002	6-11 13 16-17 19 24-25 27 31 33-35
fetal liver-spleen	Columbia University	FLS003	12-13
fetal liver	Invitrogen	FLV001	7-8 12 31 33
fetal liver	Clontech	FLV004	12 32
fetal muscle	Invitrogen	FMS001	3 5 10 27 31
fetal muscle	Invitrogen	FMS002	18
fetal skin	Invitrogen	FSK001	5 10 16-17 24 27-28 30-33
fetal skin	Invitrogen	FSK002	16-17
umbilical cord	BioChain	FUC001	4 7-8 13 23 28 31-32 34-35
fetal brain	GIBCO	HFB001	1-2 7-8 15 18 23-24 27 30
infant brain	Columbia University	IB2002	6-8 15 18-19 23 27 30-31 33
infant brain	Columbia University	IB2003	2 6-8 30
infant brain	Columbia University	IBS001	33
lung, fibroblast	Stratagene	LFB001	7-8 29 31
lung tumor	Invitrogen	LGT002	4 7-8 11 15-20 23-25 28 30- 31 34-35
lymphocytes	ATCC	LPC001	6 9 28-29
leukocyte	GIBCO	LUC001	1-2 4 6 10 19-20 24-25 28-30
leukocyte	Clontech	LUC003	24 28 35
Melanoma from cell line ATCC CRL # 1424	Clontech	MEL004	24 31 33
mammary gland	Invitrogen	MMG001	5-8 10 12 16-17 23 28 30-32 35
induced neuron cells	Stratagene	NTD001	1 14 29
retinoid acid induced neuronal cells	Stratagene	NTR001	11
neuronal cells	Stratagene	NTU001	14 31 33
prostate	Clontech	PRT001	23 35
rectum	Invitrogen	REC001	31
salivary gland	Clontech	SAL001	1 7-8 13 23
small intestine	Clontech	SIN001	6 16-18 23 27 31-32 35
skeletal muscle	Clontech	SKM001	16-18
skeletal muscle	Clontech	SKMs03	16-17
spinal cord	Clontech	SPC001	1 20 23 31
adult spleen	Clontech	SPLc01	6 13 16-17
thalamus	Clontech	THA002	7-9 21
thymus	Clontech	THM001	4 23 33 35
thymus	Clontech	THMc02	4 6-8 11 30 32-34
thyroid gland	Clontech	THR001	4 7-8 10 23-25 27-28 31 35
trachea	Clontech	TRC001	29
uterus	Clontech	UTR001	5 31

TABLE 2

SEQ ID NO:	CORRESPONDING SEQ ID NO. IN U.S.S.N 09/577,409	ACCESSION NUMBER	DESCRIPTION	SMITH-WATERMAN SCORE	% IDENTITY
1	360	U60315	Molluscum contagiosum virus subtype 1 MC015L	76	30
2	1175	A17783	unidentified NC28	583	100
3	3964	AF190819	Homo sapiens muscle beta 1 intergrin cytoplasmic domain binding protein MIBP	568	96
4	5522	D00762	Homo sapiens proteasome subunit C8	686	93
5	5537	X76091	Homo sapiens DNA binding protein RFX2	3747	99
6	5573	AK000004	Homo sapiens FLJ00004 protein	3106	96
7	5868	U77718	Homo sapiens pinin	89	20
8	5868	M87306	Tetrahymena thermophila micronuclear linker histone polyprotein	130	20
9	6265	U12329	Cricetulus griseus mutant sterol regulatory element binding protein-2	373	94
10	7013	U80735	Homo sapiens CAGF28	3779	97
11	7017	AC004235	Homo sapiens Myt1	69	50
12	7019	X00129	Homo sapiens precursor RBP	1005	96
13	8583	X97675	Homo sapiens plakophilin 2a	4288	99
14	8792	J02459	bacteriophage lambda I (tail component;223)	1111	99
15	8816	U95825	Homo sapiens androgen-induced prostate proliferative shutoff associated protein	828	98
16	8916	X95325	Homo sapiens DNA-binding protein	1975	100
17	8916	AF171061	Canis familiaris Y-box protein ZONAB-A	1426	89
18	8950	X99584	Homo sapiens SMT3A protein	522	98
19	9067	U37429	Caenorhabditis elegans similar to M. musculus MER5 and other AHPC/TSA proteins	902	47

TABLE 2

SEQ ID NO:	CORRESPONDING SEQ ID NO. IN U.S.S.N 09/577,409	ACCESSION NUMBER	DESCRIPTION	SMITH-WATERMAN SCORE	% IDENTITY
20	9254	X78926	Homo sapiens zinc finger protein	426	50
21	9285	J02459	bacteriophage lambda E (capsid component;341)	1754	100
22	10049	AF116695	Homo sapiens PRO2221	159	48
23	10858	AF157317	Homo sapiens AD-015 protein	869	100
24	11054	AF090931	Homo sapiens PRO0483	107	46
25	11872	AJ009698	Rattus norvegicus embigin protein	1080	65
26	12230	AF124512	Homo sapiens BVES	1763	100
27	12369	M15888	Bos taurus endozepine-related protein precursor	372	40
28	12464	Z99259	Schizosaccharomyces pombe putative phosphotransferase	396	43
29	12708	U28831	Homo sapiens protein that is immuno-reactive with anti-PTH polyclonal antibodies	518	55
30	13024	X58833	Streptomyces coelicolor A3(2) actVA 4	271	31
31	13199	X04412	Homo sapiens plasma gelsolin	4101	100
32	13601	X13482	Homo sapiens U2 snRNP-specific A' protein (AA 1-255)	1284	99
33	13666	X59618	Homo sapiens small subunit ribonucleotide reductase	1389	100
34	13749	AC004882	Homo sapiens similar to cytochrome Bc1 J chain; similar to 1BGY (PID:g4139401)	326	100
35	14042	AL034374	Homo sapiens dJ483K16.1.1 (novel protein (isoform 1))	1651	100

TABLE 3

SEQ ID NO:	ACCESSION NO.	DESCRIPTION	RESULTS*
1	PR00651	5-HYDROXYTRYPTAMINE 2B RECEPTOR SIGNATURE	PR00651H 5.59 7.857e-07 178-201
2	PR00437	SMALL CXC CYTOKINE FAMILY SIGNATURE	PR00437C 14.85 9.640e-10 79-98
3	PR00364	DISEASE RESISTANCE PROTEIN SIGNATURE	PR00364A 8.19 9.080e-09 4-20
4	PF00227	Proteasome A-type and B-type.	PF00227 14.68 6.294e-10 15-27
5	PD02699	PROTEIN DNA-BINDING BINDING DNA.	PD02699C 24.84 1.000e-40 614-661 PD02699A 8.91 3.250e-35 235-264 PD02699B 18.28 6.571e-21 500-524
6	BL00741	Guanine-nucleotide dissociation stimulators CDC24 family sign.	BL00741B 14.27 3.769e-09 289-312
7	PR00624	HISTONE H5 SIGNATURE	PR00624G 4.08 6.900e-09 180-200
9	BL00038	Myc-type, 'helix-loop-helix' dimerization domain proteins.	BL00038B 16.97 9.027e-10 57-78
10	BL00795	Involucrin proteins.	BL00795C 17.06 1.105e-10 362-407 BL00795C 17.06 6.651e-10 377-422 BL00795C 17.06 6.965e-10 360-405 BL00795C 17.06 7.698e-10 388-433 BL00795C 17.06 2.900e-09 374-419 BL00795C 17.06 3.800e-09 361-406 BL00795C 17.06 5.200e-09 391-436 BL00795C 17.06 9.200e-09 390-435
11	PR00003	4-DISULPHIDE CORE SIGNATURE	PR00003C 7.69 7.545e-06 13-23
12	BL00213	Lipocalin proteins.	BL00213A 12.95 3.368e-10 40-54 BL00213B 8.78 9.500e-10 128-139
15	PR00929	AT-HOOK-LIKE DOMAIN SIGNATURE	PR00929C 5.26 8.759e-09 55-66
16	BL00048	Protamine P1 proteins.	BL00048 6.39 2.688e-09 168-195
17	BL00048	Protamine P1 proteins.	BL00048 6.39 7.750e-09 168-195
18	BL00299	Ubiquitin domain proteins.	BL00299 28.84 2.250e-25 32-84
19	PF00534	Glycosyl transferases group 1.	PF00534B 14.47 8.364e-15 325-349

\* Results include in order: accession number subtype; raw score; p-value; position of signature in amino acid sequence.

TABLE 3

SEQ ID NO:	ACCESSION NO.	DESCRIPTION	RESULTS*
20	PR00048	C2H2-TYPE ZINC FINGER SIGNATURE	PR00048A 10.52 9.526e-11 153-167 PR00048A 10.52 1.000e-09 181-195 PR00048B 6.02 6.684e-09 141-151
21	PF00546	7S seed storage protein.	PF00546G 20.09 9.443e-06 191-227
22	PR00402	TEC/BTK DOMAIN SIGNATURE	PR00402A 16.09 9.556e-06 40-60
24	PR00387	3'5'-CYCLIC NUCLEOTIDE PHOSPHODIESTERASE SIGNATURE	PR00387D 10.81 4.778e-06 244-261
26	PF00603	Influenza RNA-dependant RNA polymerase subunit PA.	PF00603A 8.25 5.836e-06 235-290
27	PR00689	ACYL-COA-BINDING PROTEIN SIGNATURE	PR00689C 6.73 2.350e-15 57-73 PR00689B 16.87 7.894e-15 34-53 PR00689D 12.35 6.318e-11 78-96 PR00689A 13.38 6.523e-10 13-29
28	PR00652	5-HYDROXYTRYPTAMINE 7 RECEPTOR SIGNATURE	PR00652G 10.94 2.500e-06 29-47
29	PR00541	MUSCARINIC M4 RECEPTOR SIGNATURE	PR00541B 8.49 5.781e-06 72-88 PR00541B 8.49 9.813e-06 174-190
30	BL00895	3-hydroxyisobutyrate dehydrogenase proteins.	BL00895A 12.61 5.280e-08 8-29
31	PR00597	GELSOLIN FAMILY SIGNATURE	PR00597G 8.55 1.429e-28 690-713 PR00597A 12.96 1.000e-25 375-397 PR00597D 12.77 4.522e-24 522-543 PR00597C 14.19 9.000e-23 489-508 PR00597H 15.32 5.500e-22 719-739 PR00597E 13.46 4.130e-21 576-597 PR00597F 16.29 4.522e-21 634-654 PR00597B 9.78 7.000e-20 464-481 PR00597A 12.96 4.575e-09 742-764 PR00597B 9.78 5.629e-09 85-102 PR00597D 12.77 7.723e-09 144-165
32	DM00315	072 RIBONUCLEASE INHIBITOR.	DM00315B 6.84 6.776e-08 88-100
33	BL00368	Ribonucleotide reductase small subunit proteins.	BL00368A 36.98 1.000e-40 84-139 BL00368B 22.06 5.846e-26 161-187
34	PR00202	ANNEXIN TYPE VI SIGNATURE	PR00202A 10.28 1.000e-05 37-46
35	PR00121	SODIUM/POTASSIUM-TRANSPORTING	PR00121G 6.89 7.525e-



Variable	Mean	SD	Min	Max
Age	34.5	10.5	18	65
Gender	0.5	0.5	0	1
Marital status	0.5	0.5	0	1
Education	12.5	1.5	9	16
Income	15.5	5.5	10	25
Occupation	1.5	1.5	0	3
Health status	1.5	1.5	0	3
Stress level	2.5	1.5	0	4
Life satisfaction	3.5	1.5	1	5
Resilience	4.5	1.5	2	6
Optimism	5.5	1.5	3	7
Gratitude	6.5	1.5	4	8
Forgiveness	7.5	1.5	5	9
Empathy	8.5	1.5	6	10
Compassion	9.5	1.5	7	11
Kindness	10.5	1.5	8	12
Generosity	11.5	1.5	9	13
Patience	12.5	1.5	10	14
Self-control	13.5	1.5	11	15
Emotional stability	14.5	1.5	12	16
Psychological well-being	15.5	1.5	13	17
Life purpose	16.5	1.5	14	18
Meaning in life	17.5	1.5	15	19
Existential fulfillment	18.5	1.5	16	20
Transcendental experience	19.5	1.5	17	21
Spiritual growth	20.5	1.5	18	22
Inner peace	21.5	1.5	19	23
Harmony with nature	22.5	1.5	20	24
Connection to the universe	23.5	1.5	21	25
Universal love	24.5	1.5	22	26
Oneness with all things	25.5	1.5	23	27
Divine spark	26.5	1.5	24	28
Sacredness of life	27.5	1.5	25	29
Reverence for creation	28.5	1.5	26	30
Worship of the divine	29.5	1.5	27	31
Devotion to God	30.5	1.5	28	32
Religious faith	31.5	1.5	29	33
Spiritual journey	32.5	1.5	30	34
Path to enlightenment	33.5	1.5	31	35
Search for truth	34.5	1.5	32	36
Quest for meaning	35.5	1.5	33	37
Exploration of self	36.5	1.5	34	38
Discovery of purpose	37.5	1.5	35	39
Realization of potential	38.5	1.5	36	40
Attainment of wisdom	39.5	1.5	37	41
Acquisition of knowledge	40.5	1.5	38	42
Development of character	41.5	1.5	39	43
Cultivation of virtues	42.5	1.5	40	44
Pursuit of excellence	43.5	1.5	41	45
Striving for perfection	44.5	1.5	42	46
Aspiration for greatness	45.5	1.5	43	47
Yearning for transcendence	46.5	1.5	44	48
Longing for the divine	47.5	1.5	45	49
Desire for union with God	48.5	1.5	46	50
Craving for spiritual fulfillment	49.5	1.5	47	51
Thirst for eternal life	50.5	1.5	48	52
Hope for resurrection	51.5	1.5	49	53
Belief in afterlife	52.5	1.5	50	54
Trust in divine plan	53.5	1.5	51	55
Acceptance of fate	54.5	1.5	52	56
Submission to God's will	55.5	1.5	53	57
Surrender to the universe	56.5	1.5	54	58
Letting go of attachments	57.5	1.5	55	59
Detachment from worldly desires	58.5	1.5	56	60
Renunciation of ego	59.5	1.5	57	61
Annihilation of self	60.5	1.5	58	62
Transcendence of individuality	61.5	1.5	59	63
Union with the divine	62.5	1.5	60	64
Ecstasy of spiritual experience	63.5	1.5	61	65
Bliss of divine love	64.5	1.5	62	66
Tranquility of inner peace	65.5	1.5	63	67
Serenity of the soul	66.5	1.5	64	68
Peace of mind	67.5	1.5	65	69
Harmony of the heart	68.5	1.5	66	70
Balance of the mind	69.5	1.5	67	71
Stability of the spirit	70.5	1.5	68	72
Strength of the will	71.5	1.5	69	73
Power of the soul	72.5	1.5	70	74
Victory over adversity	73.5	1.5	71	75

TABLE 4

SEQ ID NO:	PFAM NAME	DESCRIPTION	p-value	PFAM SCORE
2	il8	Small cytokines (intecrine/chemokine), inter	3.6e-38	131.6
4	proteasome	Proteasome A-type and B-type	6.1e-39	142.8
6	FYVE	FYVE zinc finger	2.8e-12	48.6
9	HLH	Helix-loop-helix DNA-binding domain	1.4e-05	31.9
10	BRCT	BRCA1 C Terminus (BRCT) domain	1.2e-28	108.6
12	lipocalin	lipocalin	2.5e-41	146.4
13	Armadillo_seg	Armadillo/beta-catenin-like repeats	7.8e-10	46.1
16	CSD	'Cold-shock' DNA-binding domain	1.1e-33	124.6
17	CSD	'Cold-shock' DNA-binding domain	1.1e-33	124.6
19	Glycos_transf_1	Glycosyl transferases group 1	1.7e-38	138.4
20	zf-C2H2	Zinc finger, C2H2 type	0.0005	26.8
25	ig	Immunoglobulin domain	1e-06	26.7
27	ACBP	Acyl CoA binding protein	1.1e-39	145.2
28	DUF60	Domain of unknown function	2.2e-36	134.3
31	Gelsolin	Gelsolin repeat.	4.3e-112	385.8
32	LRR	Leucine Rich Repeat	0.00025	27.8
33	ribonuc_red	Ribonucleotide reductases	1.5e-122	358.6
35	GNS1_SUR4	GNS1/SUR4 family	2.7e-09	-26.6

TABLE 5

SEQ ID NO:	POSITION OF SIGNAL IN AMINO ACID SEQUENCE	maxS (MAXIMUM SCORE)	meanS (MEAN SCORE)
2	1-33	0.976	0.754
10	1-38	0.901	0.607
12	1-22	0.929	0.838
25	1-32	0.949	0.704
31	1-27	0.966	0.909
34	1-42	0.895	0.609

## CLAIMS

WHAT IS CLAIMED IS:

1. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 – 35, a mature protein coding portion of SEQ ID NO: 1 – 35, an active domain of SEQ ID NO: 1 – 35, and complementary sequences thereof.
2. An isolated polynucleotide encoding a polypeptide with biological activity, wherein said polynucleotide hybridizes to the polynucleotide of claim 1 under stringent hybridization conditions.
3. An isolated polynucleotide encoding a polypeptide with biological activity, wherein said polynucleotide has greater than about 90% sequence identity with the polynucleotide of claim 1.
4. The polynucleotide of claim 1 wherein said polynucleotide is DNA.
5. An isolated polynucleotide of claim 1 wherein said polynucleotide comprises the complementary sequences.
6. A vector comprising the polynucleotide of claim 1.
7. An expression vector comprising the polynucleotide of claim 1.
8. A host cell genetically engineered to comprise the polynucleotide of claim 1.
9. A host cell genetically engineered to comprise the polynucleotide of claim 1 operatively associated with a regulatory sequence that modulates expression of the polynucleotide in the host cell.
10. An isolated polypeptide, wherein the polypeptide is selected from the group consisting of:
  - (a) a polypeptide encoded by any one of the polynucleotides of claim 1; and

(b) a polypeptide encoded by a polynucleotide hybridizing under stringent conditions with any one of SEQ ID NO: 1 – 35.

11. A composition comprising the polypeptide of claim 10 and a carrier.

12. An antibody directed against the polypeptide of claim 10.

13. A method for detecting the polynucleotide of claim 1 in a sample, comprising:

a) contacting the sample with a compound that binds to and forms a complex with the polynucleotide of claim 1 for a period sufficient to form the complex; and

b) detecting the complex, so that if a complex is detected, the polynucleotide of claim 1 is detected.

14. A method for detecting the polynucleotide of claim 1 in a sample, comprising:

a) contacting the sample under stringent hybridization conditions with nucleic acid primers that anneal to the polynucleotide of claim 1 under such conditions;

b) amplifying a product comprising at least a portion of the polynucleotide of claim 1; and

c) detecting said product and thereby the polynucleotide of claim 1 in the sample.

15. The method of claim 14, wherein the polynucleotide is an RNA molecule and the method further comprises reverse transcribing an annealed RNA molecule into a cDNA polynucleotide.

16. A method for detecting the polypeptide of claim 10 in a sample, comprising:

a) contacting the sample with a compound that binds to and forms a complex with the polypeptide under conditions and for a period sufficient to form the complex; and

b) detecting formation of the complex, so that if a complex formation is detected, the polypeptide of claim 10 is detected.

17. A method for identifying a compound that binds to the polypeptide of claim 10, comprising:

a) contacting the compound with the polypeptide of claim 10 under conditions sufficient to form a polypeptide/compound complex; and

b) detecting the complex, so that if the polypeptide/compound complex is detected, a compound that binds to the polypeptide of claim 10 is identified.

18. A method for identifying a compound that binds to the polypeptide of claim 10, comprising:

a) contacting the compound with the polypeptide of claim 10, in a cell, under conditions sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a reporter gene sequence in the cell; and

b) detecting the complex by detecting reporter gene sequence expression, so that if the polypeptide/compound complex is detected, a compound that binds to the polypeptide of claim 10 is identified.

19. A method of producing the polypeptide of claim 10, comprising,

a) culturing a host cell comprising a polynucleotide sequence selected from the group consisting of a polynucleotide sequence of SEQ ID NO: 1-35, a mature protein coding portion of SEQ ID NO: 1-35, an active domain of SEQ ID NO: 1-35, complementary sequences thereof and a polynucleotide sequence hybridizing under stringent conditions to SEQ ID NO: 1-35, under conditions sufficient to express the polypeptide in said cell; and

b) isolating the polypeptide from the cell culture or cells of step (a).

20. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of any one of the polypeptides from the Sequence Listing, the mature protein portion thereof, or the active domain thereof.

21. The polypeptide of claim 20 wherein the polypeptide is provided on a polypeptide array.

23. The collection of claim 22, wherein the collection is provided on a nucleic acid array.

24. The collection of claim 23, wherein the array detects full-matches to any one of the polynucleotides in the collection.

25. The collection of claim 23, wherein the array detects mismatches to any one of the polynucleotides in the collection.

26. The collection of claim 22, wherein the collection is provided in a computer-readable format.

27. A method of treatment comprising administering to a mammalian subject in need thereof a therapeutic amount of a composition comprising a polypeptide of claim 10 or 20 and a pharmaceutically acceptable carrier.

28. A method of treatment comprising administering to a mammalian subject in need thereof a therapeutic amount of a composition comprising an antibody that specifically binds to a polypeptide of claim 10 or 20 and a pharmaceutically acceptable carrier.

[illegible]



## DECLARATION FOR PATENT APPLICATION AND POWER OF ATTORNEY

As [a] below named inventor(s), I/we hereby declare that:

**Y. Tom Tang, Chenghua Liu, Ping Zhou, Vinod Asundi, Jie Zhang, Qing A. Zhao,  
Feiyan Ren, Aidong J. Xue, Yonghong Yang, Tom Wehrman, Radoje T. Drmanac**

My/our residence, post office address and citizenship is/are as stated below next to my/our name(s).

I/we believe I/we am/are an/the original, first and sole/joint inventor of the subject matter which is claimed and for which a patent is sought on the invention entitled: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES, the specification of which

  X   is attached hereto.

       was filed on [date] as Application Serial Number [            ]  
and was amended on [date].

I/We hereby state that I/we have reviewed and understand the contents of the above-identified specification, including the claims as amended by any amendment referred to above.

I/We acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, Section 1.56(a).

I/We hereby claim foreign priority benefits under Title 35, United States Code, § 119 of any foreign application(s) for patent or inventor's certificate, listed below and so identified, and I/we have also identified below any foreign application for patent or inventor's certificate on this invention filed by me or my legal representatives or assigns and having a filing date before that of the application on which priority is claimed:

NUMBER	COUNTRY	DAY/MONTH/ YEAR FILED	PRIORITY CLAIMED - YES OR NO

I/We hereby claim the benefit under Title 35, United States Code, § 120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, § 112, I/we acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, § 1.56(a) which occurred between the filing date of the prior application and the national or PCT international filing date of this application.

SERIAL NUMBER	FILING DATE	STATUS
09/577,409	May 18, 2000	Pending
09/515,126	February 28, 2000	Pending

I/We hereby declare that all statements made herein of my/our own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under § 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

I/We hereby appoint the following attorneys and agents, with full power of substitution and revocation, to prosecute this application and to transact all business in the United States Patent and Trademark Office connected therewith and request that all correspondence and telephone calls with respect to this application be directed to Leslie A. Mooi, HYSEQ, INC., 670 Almanor Avenue, Sunnyvale, CA 94085, Telephone No. (408) 524-8100:

ATTORNEY	REGISTRATION NO.
Petrina S. Hsi	38,496
Leslie A. Mooi	37,047

Full name of  
first joint inventor: Y. Tom Tang

Inventor's signature: \_\_\_\_\_ Date: \_\_\_\_\_

Residence and Post  
Office Address: 4230 Ranwick Court, San Jose, CA 95118

Citizenship: United States of America

Full name of  
second joint inventor: Chenghua Liu

Inventor's signature: \_\_\_\_\_ Date: \_\_\_\_\_

Residence and Post  
Office Address: 1125 Ranchero Way, Apt. # 14, San Jose, CA 95117

Citizenship: People's Republic of China

Case No. T-4349360

Full name of  
third joint inventor: Ping Zhou

Inventor's signature: \_\_\_\_\_ Date: \_\_\_\_\_

Residence and Post  
Office Address: 1461 Japaur Lane, San Jose, CA 95132

Citizenship: People's Republic of China

Full name of  
fourth joint inventor: Vinod Asundi

Inventor's signature: \_\_\_\_\_ Date: \_\_\_\_\_

Residence and Post  
Office Address: 709 Foster City Blvd., Foster City, CA 94404

Citizenship: United States of America

Full name of  
fifth joint inventor: Jie Zhang

Inventor's signature: \_\_\_\_\_ Date: \_\_\_\_\_

Residence and Post  
Office Address: 20800 Homestead Road, #38B, Cupertino, CA 95014

Citizenship: People's Republic of China

Full name of  
sixth joint inventor: Qing A. Zhao

Inventor's signature: \_\_\_\_\_ Date: \_\_\_\_\_

Residence and Post  
Office Address: 1028 S. de Anza Blvd., Apt. B-210, San Jose, CA 95129

Citizenship: People's Republic of China

Case No. 788CIP2C

Full name of seventh joint inventor: Feiyan Ren

Inventor's signature: \_\_\_\_\_ Date: \_\_\_\_\_

Residence and Post Office Address: 20685 Garden Manor Court, Cupertino, CA 95014

Citizenship: People's Republic of China

Full name of eighth joint inventor: Aidong J. Xue

Inventor's signature: \_\_\_\_\_ Date: \_\_\_\_\_

Residence and Post Office Address: 1621 S. Mary Avenue, Sunnyvale, CA 94087

Citizenship: People's Republic of China

Full name of ninth joint inventor: Yonghong Yang

Inventor's signature: \_\_\_\_\_ Date: \_\_\_\_\_

Residence and Post Office Address: 4230 Ranwick Court, San Jose, CA 95118

Citizenship: People's Republic of China

Full name of tenth joint inventor: Tom Wehrman

Inventor's signature: \_\_\_\_\_ Date: \_\_\_\_\_

Residence and Post Office Address: 300 Pasteur Drive, Edwards R314, Stanford University Medical Center, Stanford, CA 94305

Citizenship: United States of America



# SEQUENCE LISTING

<110> Tang, Y. Tom  
 Liu, Chenghua  
 Zhou, Ping  
 Asundi, Vinod  
 Zhang, Jie  
 Zhao, Qing A.  
 Ren, Feiyan  
 Xue, Aidong J.  
 Yang, Yonghong  
 Wehrman, Tom  
 Drmanac, Radoje T.

<120> Novel Nucleic Acids and  
 Polypeptides

<130> 788CIP2C

<140> To Be Assigned  
 <141>

<150> 09/577,409  
 <151> 2000-05-18

<150> 09/515,126  
 <151> 2000-02-28

<160> 35

<170> pt\_FL\_genes Version 2.0

<210> 1  
 <211> 941  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> (133)..(786)

<400> 1  
 tcgcatgcac gcgtacgtaa gcttggatcc tctagagcgg ccgcctggct gctcctcggg 60  
 gcggcgaccg ggctcacgcg cgggcccgcc gcggccttca ccgccgcgcg ctctgacgcc 120

```

ggcataaggg cc      atg tgt tct gaa att att ttg agg caa gaa gtt ttg      168
                    Met Cys Ser Glu Ile Ile Leu Arg Gln Glu Val Leu
                      1             5             10

aaa gat ggt ttc cac aga gac ctt tta atc aaa gtg aag ttt ggg gaa      216
Lys Asp Gly Phe His Arg Asp Leu Leu Ile Lys Val Lys Phe Gly Glu
                      15             20             25

agc att gag gac ttg cac acc tgc cgt ctc tta att aaa cag gac att      264
Ser Ile Glu Asp Leu His Thr Cys Arg Leu Leu Ile Lys Gln Asp Ile
                      30             35             40

cct gca gga ctt tat gtg gat ccg tat gag ttg gct tca tta cga gag      312
Pro Ala Gly Leu Tyr Val Asp Pro Tyr Glu Leu Ala Ser Leu Arg Glu
                      45             50             55             60

aga aac ata aca gag gca gtg atg gtt tca gaa aat ttt gat ata gag      360
Arg Asn Ile Thr Glu Ala Val Met Val Ser Glu Asn Phe Asp Ile Glu
                      65             70             75

gcc cct aac tat ttg tcc aag gag tct gaa gtt ctc att tat gcc aga      408
Ala Pro Asn Tyr Leu Ser Lys Glu Ser Glu Val Leu Ile Tyr Ala Arg
                      80             85             90

cga gat tca cag tgc att gac tgt ttt caa gcc ttt ttg cct gtg cac      456
Arg Asp Ser Gln Cys Ile Asp Cys Phe Gln Ala Phe Leu Pro Val His
                      95             100             105

tgc cgc tat cat cgg ccg cac agt gaa gat gga gaa gcc tcg att gtg      504
Cys Arg Tyr His Arg Pro His Ser Glu Asp Gly Glu Ala Ser Ile Val
                      110             115             120

gtc aat aac cca gat ttg ttg atg ttt tgt gac caa gag ttc ccg att      552
Val Asn Asn Pro Asp Leu Leu Met Phe Cys Asp Gln Glu Phe Pro Ile
                      125             130             135             140

ttg aaa tgc tgg gct cac tca gaa gtg gca gcc cct tgt gct ttg gag      600
Leu Lys Cys Trp Ala His Ser Glu Val Ala Ala Pro Cys Ala Leu Glu
                      145             150             155

aat gag gat atc tgc caa tgg aac aag atg aag tat aaa tca gta tat      648
Asn Glu Asp Ile Cys Gln Trp Asn Lys Met Lys Tyr Lys Ser Val Tyr
                      160             165             170

aag aat gtg att cta caa gtt cca gtg gga ctg act gta cat acc tct      696
Lys Asn Val Ile Leu Gln Val Pro Val Gly Leu Thr Val His Thr Ser
                      175             180             185

cta gta tgt tct gtg act ctg ctc att aca atc ctg tgc tct aca ttg      744
Leu Val Cys Ser Val Thr Leu Leu Ile Thr Ile Leu Cys Ser Thr Leu
                      190             195             200

atc ctt gta gca gtt ttc aaa tat ggc cat ttt tcc cta taa gttttat      793
Ile Leu Val Ala Val Phe Lys Tyr Gly His Phe Ser Leu *
                      205             210             215

gtagttaaat gcttcctaga aacctaaata agatctatta atttctgacg agaggtgttc      853

```





Ser	Pro	Gln	Gly	Leu	Ala	Gln	Pro	Val	Gly	Ile	Asn	Thr	Ser	Thr	Thr		
	30						35					40					
tgc	tgc	tac	aga	ttt	atc	aat	aag	aaa	atc	cct	aag	cag	agg	ctg	gag	1096	
Cys	Cys	Tyr	Arg	Phe	Ile	Asn	Lys	Lys	Ile	Pro	Lys	Gln	Arg	Leu	Glu		
	45						50				55						
agc	tac	aga	agg	acc	acc	agt	agc	cac	tgt	ccc	cgg	gaa	gct	gta	atc	1144	
Ser	Tyr	Arg	Arg	Thr	Thr	Ser	Ser	His	Cys	Pro	Arg	Glu	Ala	Val	Ile		
	60				65					70				75			
ttc	aag	acc	aaa	ctg	gac	aag	gag	atc	tgt	gct	gac	ccc	aca	cag	aag	1192	
Phe	Lys	Thr	Lys	Leu	Asp	Lys	Glu	Ile	Cys	Ala	Asp	Pro	Thr	Gln	Lys		
				80					85					90			
tgg	gtc	cag	gac	ttt	atg	aag	cac	ctg	gac	aag	aaa	acc	caa	act	cca	1240	
Trp	Val	Gln	Asp	Phe	Met	Lys	His	Leu	Asp	Lys	Lys	Thr	Gln	Thr	Pro		
			95					100					105				
aag	ctt	tga	acattca	tgactgaact	gaaaacaagc	catgacttga	gaaacaaata									1296	
Lys	Leu	*															
			110														
attt	gtatac	cctgtccttt	ctcagagtgg	ttctgagatt	attttaatct	aattctaagg										1356	
aatat	gagct	ttatgtaata	atgtgaatca	tggtttttct	tagtagattt	taaaagttat										1416	
taat	attttta	attttaatctt	ccatggattt	tggtgggttt	tgaacataaa	gccttggatg										1476	
tatat	gtcat	ctcagtgtctg	taaaaactgt	gggatgtctc	tcccttctct	acctcatggg										1536	
ggat	tattgtat	aagtccttgc	agaatcagt	gcaaagattt	gctttaattg	ttaagatatg										1596	
atgt	ccctat	ggaagcatat	tgttattata	taattacata	tttgcatatg	tatgactccc										1656	
aaat	tttcac	ataaaataga	tttttgtata	acaaaaacaa	aaaaaaaaaa											1706	

<210> 3  
 <211> 1001  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> (227) .. (847)

<400> 3																	
attgatcgcc	tgcggtaccg	gtccggaatt	cccgggtcga	cccacgcgtc	cgcgccctccg											60	
ccccatcccc	aggggccgcc	tccccgggg	cggcctccag	gctgccgaga	cctataaagg											120	
cgccaggttt	tctcaatgaa	gccgggacgc	actccggagc	gcactgcgtg	gtcgcaccct											180	
accggggctg	ccttgggaagt	cgtccccgcc	gcccctccgc	accggc	atg aag ctc											235	

																		Met Lys Leu	
																		1	
atc	gtg	ggc	atc	gga	ggc	atg	acc	aac	ggc	ggc	aag	acc	acg	ctg	acc		283		
Ile	Val	Gly	Ile	Gly	Gly	Met	Thr	Asn	Gly	Gly	Lys	Thr	Thr	Leu	Thr				
	5					10					15								
aac	agc	ctg	ctc	aga	gcc	ctg	ccc	aac	tgc	tgc	gtg	atc	cat	cag	gat		331		
Asn	Ser	Leu	Leu	Arg	Ala	Leu	Pro	Asn	Cys	Cys	Val	Ile	His	Gln	Asp				
	20				25					30					35				
gac	ttc	ttc	aag	ccc	caa	gac	caa	ata	gca	gtt	ggg	gaa	gac	ggc	ttc		379		
Asp	Phe	Phe	Lys	Pro	Gln	Asp	Gln	Ile	Ala	Val	Gly	Glu	Asp	Gly	Phe				
				40					45					50					
aaa	cag	tgg	gac	gtg	ctg	gag	tct	ctg	gac	atg	gag	gcc	atg	ctg	gac		427		
Lys	Gln	Trp	Asp	Val	Leu	Glu	Ser	Leu	Asp	Met	Glu	Ala	Met	Leu	Asp				
			55					60					65						
acc	gtg	cag	gcc	tgg	ctg	agc	agc	ccg	cag	aag	ttt	gcc	cgt	gcc	cac		475		
Thr	Val	Gln	Ala	Trp	Leu	Ser	Ser	Pro	Gln	Lys	Phe	Ala	Arg	Ala	His				
			70					75				80							
ggg	gtc	agc	gtc	cag	cca	gag	gcc	tcg	gac	acc	cac	atc	ctc	ctt	ctg		523		
Gly	Val	Ser	Val	Gln	Pro	Glu	Ala	Ser	Asp	Thr	His	Ile	Leu	Leu	Leu				
	85					90					95								
gaa	ggc	ttc	ctg	ctc	tac	agc	tac	aat	acc	cgc	aac	tac	aca	gtc	cct		571		
Glu	Gly	Phe	Leu	Leu	Tyr	Ser	Tyr	Asn	Thr	Arg	Asn	Tyr	Thr	Val	Pro				
	100				105					110					115				
gat	ccc	ccc	ggc	ctc	ttc	gat	ggc	cac	gtg	tgg	ccc	atg	tac	cag	aag		619		
Asp	Pro	Pro	Gly	Leu	Phe	Asp	Gly	His	Val	Trp	Pro	Met	Tyr	Gln	Lys				
				120					125					130					
tat	agg	cag	gag	atg	gag	gcc	aac	ggt	gtg	gaa	gtg	gtc	tac	ctg	gac		667		
Tyr	Arg	Gln	Glu	Met	Glu	Ala	Asn	Gly	Val	Glu	Val	Val	Tyr	Leu	Asp				
			135					140					145						
ggc	atg	aag	tcc	cga	gag	gag	ctc	ttc	cgt	gaa	gtc	ctg	gaa	gac	att		715		
Gly	Met	Lys	Ser	Arg	Glu	Glu	Leu	Phe	Arg	Glu	Val	Leu	Glu	Asp	Ile				
		150					155					160							
cag	aac	tcg	ctg	ctg	aac	cgc	tcc	cag	gaa	tca	gcc	ccc	tcc	ccg	gct		763		
Gln	Asn	Ser	Leu	Leu	Asn	Arg	Ser	Gln	Glu	Ser	Ala	Pro	Ser	Pro	Ala				
	165					170					175								
cgc	cca	gcc	agg	aca	cag	gga	ccc	gga	cgc	gga	tgc	ggc	cac	aga	acg		811		
Arg	Pro	Ala	Arg	Thr	Gln	Gly	Pro	Gly	Arg	Gly	Cys	Gly	His	Arg	Thr				
	180				185					190					195				
gcc	agg	cct	gca	gcg	tcc	cag	cag	gac	agc	atg	tga	gcgt	ttccctatgg				861		
Ala	Arg	Pro	Ala	Ala	Ser	Gln	Gln	Asp	Ser	Met	*								
				200				205											
gggtgtctgt	acgtaggaga	gtggaggccc	cactcccagt	tgggcgtccc	ggagctcagg												921		





ccc aga gtt cag cag gta ccc cag cag gtg cag ccg gtg cag cac gtg Pro Arg Val Gln Gln Val Pro Gln Gln Val Gln Pro Val Gln His Val 55 60 65	549
tat cct gcc cag gtg cag tac gtg gaa ggg gga gac gcc gtc tac acc Tyr Pro Ala Gln Val Gln Tyr Val Glu Gly Gly Asp Ala Val Tyr Thr 70 75 80	597
aat gga gcc ata cga aca gcc tac acc tac aac ccc gag cct cag atg Asn Gly Ala Ile Arg Thr Ala Tyr Thr Tyr Asn Pro Glu Pro Gln Met 85 90 95	645
tac gcc ccc agc agc acg gct tct tac ttc gag gcc cca ggc ggt gcc Tyr Ala Pro Ser Ser Thr Ala Ser Tyr Phe Glu Ala Pro Gly Gly Ala 100 105 110 115	693
cag gtg acc gtg gca gcc tcg tcc ccg cca gcg gtc ccc tcc cac agc Gln Val Thr Val Ala Ala Ser Ser Pro Pro Ala Val Pro Ser His Ser 120 125 130	741
atg gtg ggc atc acc atg gat gtc ggg ggg agc ccc atc gtc tcc agc Met Val Gly Ile Thr Met Asp Val Gly Gly Ser Pro Ile Val Ser Ser 135 140 145	789
gcg gga gcc tat ctc atc cac ggg ggg atg gac agc acc aga cac tcc Ala Gly Ala Tyr Leu Ile His Gly Gly Met Asp Ser Thr Arg His Ser 150 155 160	837
ctg gcc cac acc tcc cgc tca tcg ccc gcc acg ctt gaa atg gcg att Leu Ala His Thr Ser Arg Ser Ser Pro Ala Thr Leu Glu Met Ala Ile 165 170 175	885
gaa aac ctc caa aaa agc gaa gga atc aca tca cac aaa agc ggt tta Glu Asn Leu Gln Lys Ser Glu Gly Ile Thr Ser His Lys Ser Gly Leu 180 185 190 195	933
ctc aac agc cat ctc cag tgg ctg ttg gat aat tat gaa aca gcg gaa Leu Asn Ser His Leu Gln Trp Leu Leu Asp Asn Tyr Glu Thr Ala Glu 200 205 210	981
ggc gtg agt ctc ccc aga agt tct ctt tac aac cac tac ctt cgg cac Gly Val Ser Leu Pro Arg Ser Ser Leu Tyr Asn His Tyr Leu Arg His 215 220 225	1029
tgc cag gag cac aag cta gac cca gtg aac gcc gcc tcc ttc ggg aaa Cys Gln Glu His Lys Leu Asp Pro Val Asn Ala Ala Ser Phe Gly Lys 230 235 240	1077
ctg atc cgt tct gtg ttt atg ggg ctg aga acg cgg cgg ctg ggc acc Leu Ile Arg Ser Val Phe Met Gly Leu Arg Thr Arg Arg Leu Gly Thr 245 250 255	1125
agg ggc aac tcg aag tac cat tac tat ggg att cgt ctg aag ccg gac Arg Gly Asn Ser Lys Tyr His Tyr Tyr Gly Ile Arg Leu Lys Pro Asp 260 265 270 275	1173
tca cca ctg aac cgg ctg cag gag gac acg cag tac atg gcc atg cgg	1221

Ser	Pro	Leu	Asn	Arg	Leu	Gln	Glu	Asp	Thr	Gln	Tyr	Met	Ala	Met	Arg		
				280					285					290			
cag	cag	ccc	atg	cac	cag	aag	ccc	agg	tac	cgg	cca	gcc	cag	aag	acg		1269
Gln	Gln	Pro	Met	His	Gln	Lys	Pro	Arg	Tyr	Arg	Pro	Ala	Gln	Lys	Thr		
			295					300					305				
gac	agc	ctc	ggg	gac	agc	ggc	tcc	cac	agc	ggc	ctg	cac	agc	act	ccg		1317
Asp	Ser	Leu	Gly	Asp	Ser	Gly	Ser	His	Ser	Gly	Leu	His	Ser	Thr	Pro		
		310					315					320					
gaa	cag	acc	atg	gcc	gtg	cag	agc	cag	cac	cac	cag	cag	tac	ata	gat		1365
Glu	Gln	Thr	Met	Ala	Val	Gln	Ser	Gln	His	His	Gln	Gln	Tyr	Ile	Asp		
	325					330					335						
gtc	tcc	cac	gtc	ttc	ccc	gag	ttc	cca	gcg	ccc	gac	ctg	ggc	agc	ttc		1413
Val	Ser	His	Val	Phe	Pro	Glu	Phe	Pro	Ala	Pro	Asp	Leu	Gly	Ser	Phe		
	340				345					350					355		
ctg	ctg	cag	gac	ggc	gtc	aca	ctg	cac	gac	gtc	aag	gcc	ctg	cag	ctg		1461
Leu	Leu	Gln	Asp	Gly	Val	Thr	Leu	His	Asp	Val	Lys	Ala	Leu	Gln	Leu		
			360						365					370			
gtg	tac	aga	cgg	cac	tgc	gag	gca	act	gta	gat	gtg	gtg	atg	aac	ctc		1509
Val	Tyr	Arg	Arg	His	Cys	Glu	Ala	Thr	Val	Asp	Val	Val	Met	Asn	Leu		
			375					380					385				
cag	ttc	cac	tac	atc	gag	aag	ctg	tgg	ctc	tcc	ttc	tgg	aac	tct	aag		1557
Gln	Phe	His	Tyr	Ile	Glu	Lys	Leu	Trp	Leu	Ser	Phe	Trp	Asn	Ser	Lys		
		390					395					400					
gcc	tcc	tcc	agc	gac	ggc	ccc	acc	tct	ctt	cct	gcc	agt	gac	gaa	gac		1605
Ala	Ser	Ser	Ser	Asp	Gly	Pro	Thr	Ser	Leu	Pro	Ala	Ser	Asp	Glu	Asp		
	405					410				415							
ccc	gag	ggc	gcc	gtc	ctg	ccc	aag	gac	aag	ctt	atc	tcc	ctg	tgt	cag		1653
Pro	Glu	Gly	Ala	Val	Leu	Pro	Lys	Asp	Lys	Leu	Ile	Ser	Leu	Cys	Gln		
	420				425					430					435		
tgc	gac	ccc	atc	ctc	agg	tgg	atg	agg	agc	tgc	gac	cac	atc	ctc	tac		1701
Cys	Asp	Pro	Ile	Leu	Arg	Trp	Met	Arg	Ser	Cys	Asp	His	Ile	Leu	Tyr		
				440				445					450				
cag	gcg	ctg	gtg	gag	att	ctc	atc	ccc	gac	gtg	ctg	agg	ccg	gtc	ccc		1749
Gln	Ala	Leu	Val	Glu	Ile	Leu	Ile	Pro	Asp	Val	Leu	Arg	Pro	Val	Pro		
			455					460					465				
agt	acc	ttg	aca	cag	gcc	atc	cgt	aac	ttt	gcc	aag	agc	ttg	gaa	ggc		1797
Ser	Thr	Leu	Thr	Gln	Ala	Ile	Arg	Asn	Phe	Ala	Lys	Ser	Leu	Glu	Gly		
		470					475					480					
tgg	ttg	aca	aat	gcc	atg	agt	gac	ttc	cca	caa	cag	gtc	atc	cag	acc		1845
Trp	Leu	Thr	Asn	Ala	Met	Ser	Asp	Phe	Pro	Gln	Gln	Val	Ile	Gln	Thr		
		485				490					495						
aag	gtg	ggc	gtc	gtc	agt	gcc	ttc	gcc	cag	acg	ctg	cgg	cgc	tac	acg		1893
Lys	Val	Gly	Val	Val	Ser	Ala	Phe	Ala	Gln	Thr	Leu	Arg	Arg	Tyr	Thr		

500	505	510	515	
tcc ctc aac cac ctg gcg cag gcg gcc cgg gcg gtg ctg cag aac acg				1941
Ser Leu Asn His Leu Ala Gln Ala Ala Arg Ala Val Leu Gln Asn Thr	520	525	530	
tcc cag atc aac cag atg ctc agc gac ctc aac cgc gtg gac ttt gcc				1989
Ser Gln Ile Asn Gln Met Leu Ser Asp Leu Asn Arg Val Asp Phe Ala	535	540	545	
aac gtg cag gag cag gcc tcg tgg gtg tgc cag tgc gag gag agt gtg				2037
Asn Val Gln Glu Gln Ala Ser Trp Val Cys Gln Cys Glu Glu Ser Val	550	555	560	
gtg cag cgg ctg gag cag gat ttc aag ctg acc ctg cag cag cag agc				2085
Val Gln Arg Leu Glu Gln Asp Phe Lys Leu Thr Leu Gln Gln Gln Ser	565	570	575	
tcc ctg gac cag tgg gcc agc tgg ctg gac agt gtg gtc acc cag gtc				2133
Ser Leu Asp Gln Trp Ala Ser Trp Leu Asp Ser Val Val Thr Gln Val	580	585	590	595
ctg aag cag cat gcc ggc agc ccc agc ttc ccc aag gcc gcc cgg cag				2181
Leu Lys Gln His Ala Gly Ser Pro Ser Phe Pro Lys Ala Ala Arg Gln	600	605	610	
ttc ttg ctg aaa tgg tcc ttt tac agc tcc atg gtg atc cgg gac ctg				2229
Phe Leu Leu Lys Trp Ser Phe Tyr Ser Ser Met Val Ile Arg Asp Leu	615	620	625	
acc ctg cgc agc gct gcc agc ttc ggc tcc ttc cac ctc atc cgc ctg				2277
Thr Leu Arg Ser Ala Ala Ser Phe Gly Ser Phe His Leu Ile Arg Leu	630	635	640	
ctc tac gac gag tac atg ttc tac ctg gtg gag cac cgc gtc gcg gag				2325
Leu Tyr Asp Glu Tyr Met Phe Tyr Leu Val Glu His Arg Val Ala Glu	645	650	655	
gcc acc gga gag acg ccg atc gct gtg atg gga gag ttc aac gat ctc				2373
Ala Thr Gly Glu Thr Pro Ile Ala Val Met Gly Glu Phe Asn Asp Leu	660	665	670	675
gcc tct ctg tcg ctg acg ctg ctc gac aaa gat gac atg ggc gat gag				2421
Ala Ser Leu Ser Leu Thr Leu Leu Asp Lys Asp Asp Met Gly Asp Glu	680	685	690	
cag cgt ggc agc gag gcg ggc cca gac gcc cgc agc ctg ggt gag ccc				2469
Gln Arg Gly Ser Glu Ala Gly Pro Asp Ala Arg Ser Leu Gly Glu Pro	695	700	705	
ctg gta aag cgg gag cgc agt gac ccc aac cac tcc ctg cag ggc atc				2517
Leu Val Lys Arg Glu Arg Ser Asp Pro Asn His Ser Leu Gln Gly Ile	710	715	720	
tag cagc cccggccggc gcctcctcga gggtccaaaa gatgccgcct gggtcactctg				2574
*				

```

ggaacctgga tttcacccggc tccaccaaata tagtgcctct tagatgatgt gatgtttact 2634
ttgactcaag cggggggtcc cgggggtcagt gttcaagaag gaaagcagtt gttgaagcta 2694
cagaagccca ggccaggggt cccactgggt gtagctgtgc tctgggtccc caggtccctg 2754
cgcagccacc agggccccac cagaccatgg gttccttatt ctggacaaac cctcctctgc 2814
tggaaccctg gactttggcc ccagtccttt taagagctga aagggtgtgt ctgagagcac 2874
tggcctctgc actcaactgg aaccaatttt caccatgatt ttaaaaatta aaaaaacctg 2934
gcagagcccc actagaaagg gacggcctct cccctcctac ctgccaccgc tggggaggca 2994
cagagctcgc cagcgtccct ggagtcctct caggaatgga aagtggagtg cccttgtctg 3054
gtagcaccat gtgggaggga cccagctggg cgcagcgcct tgtggccttt tagatccaga 3114
cctccctgcc ggatgccccg aggcgggagg cgggtgtgtc tgcaggaacc catctccaga 3174
tgccaaagga cttgaggggc agctgacaat cgctgtgtcc cggcagatcc gcagctcgaa 3234
aaagaacaag ccacagaaac gggctcgctc gtgccaggac acagcagtgt ctttcaaaaa 3294
atcaaaacca gaagttttat cagcagcagg aaggatgtgg gactctgtcc aagttcaccg 3354
tcaccatcaa gccactgggt gtggaaggag tttgtccaac agggtcagtg tcacagccac 3414
aacttcagaa agcagccatc ccgcgtgtcg tccaaacagc agtgctgtcg tcccgctcca 3474
cggagtcttc cagaacacct ccctctgaac taatcccga gcacccacc cccagcgagt 3534
cccacctgga cctcttcaaa gtcaaaactc ttctctcggg acaaacaaat attcgggact 3594
ttcgaagccc tgaacttgag acaccaagtt tgtctccatg gaaacacgta gccgctgcag 3654
ttggacagct gcgtgctgga gttctgtccc caaaccaact gcagagaaga aagacgagcc 3714
tcggcccagg acgcgatttc taacgaaatg caatagtttt ctcttttttg aaaaatgttt 3774
ataccaatag ttttgtgcgc taattgttat tagaacgttt cttatgaatg tgtaaaaaaa 3834
aaaaaa 3840

```

```

<210> 6
<211> 3321
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> CDS
<222> (568) .. (2745)

```



<400> 6  
attattggaa cgacgacttc gtgcagaagc aggaggatgg tggttgtggg tcctctgtgt 60  
tcagccctgt taacctttaa accagcacgg ccgccgcagc tgtgagccta gcacctgatc 120  
agtggagagc tgtggattgc atgtttgttt gccattgccc ccgccaccct gcaagttgca 180  
ccttctagaa tcagcaagcc aagctcctct caccagcgt aatgatgcgg aaatgcaa 240  
gcaccatcat gttgtgaccc atattgcgaa aattagaaaa aaggaagttg tgtttcgcta 300  
ttgcacgaag ttcagcccag aggagaaact cgctgcctt cagaagacag tacctcctaa 360  
atggctctac tttgaacctg ctgggcaagg aagagatttt caaggaaacc atctaccgtg 420  
tgcaagctcc tgccggccaa cccagaccc cagcacggag ccaggcgcct gtgcccgcg 480  
accctcagca tcctcctcag aaaggctggt ggcatcagga agccctggc cagcctccac 540  
ctgagcccag tgagctcagc tttaagg atg gag tca ggc agg ggg tcc tca 591  
Met Glu Ser Gly Arg Gly Ser Ser  
1 5  
acc cct cca gga ccc att gct gcc cta ggg atg cca gac act ggg cct 639  
Thr Pro Pro Gly Pro Ile Ala Ala Leu Gly Met Pro Asp Thr Gly Pro  
10 15 20  
ggc agt tcc tcc cta ggg aag ctt cag gcg ctc cct gtt ggg ccc aga 687  
Gly Ser Ser Ser Leu Gly Lys Leu Gln Ala Leu Pro Val Gly Pro Arg  
25 30 35 40  
gcc cac tgt ggg gac cct gtc agc ctg gct gca gca ggg gac ggc tct 735  
Ala His Cys Gly Asp Pro Val Ser Leu Ala Ala Ala Gly Asp Gly Ser  
45 50 55  
cca gac ata ggc ccc acg gga gag ctg agt ggt agc tta aag atc ccc 783  
Pro Asp Ile Gly Pro Thr Gly Glu Leu Ser Gly Ser Leu Lys Ile Pro  
60 65 70  
aac cgg gac agc ggg atc gac agt ccc tcc tcc agt gtg gct gga gag 831  
Asn Arg Asp Ser Gly Ile Asp Ser Pro Ser Ser Ser Val Ala Gly Glu  
75 80 85  
aac ttt ccc tgc gag gag ggc ttg gag gct ggc cca agc ccc act gta 879  
Asn Phe Pro Cys Glu Glu Gly Leu Glu Ala Gly Pro Ser Pro Thr Val  
90 95 100  
ctg ggg gcg cac gca gag atg gcc ctg gac agc cag gtc ccg aag gtc 927  
Leu Gly Ala His Ala Glu Met Ala Leu Asp Ser Gln Val Pro Lys Val  
105 110 115 120  
acc ccc cag gag gag gcg gac agc gac gtg ggt gag gaa cct gac tct 975  
Thr Pro Gln Glu Glu Ala Asp Ser Asp Val Gly Glu Glu Pro Asp Ser  
125 130 135  
gag aac acc ccc cag aag gct gac aag gat gcc ggc ctg gcc cag cac 1023  
Glu Asn Thr Pro Gln Lys Ala Asp Lys Asp Ala Gly Leu Ala Gln His

140	145	150	
tct ggc ccc cag aag ctt ctc cac att gcc cag gag ctc ctg cac acc			1071
Ser Gly Pro Gln Lys Leu Leu His Ile Ala Gln Glu Leu Leu His Thr			
155	160	165	
gag gag acc tat gtg aag cgg ctg cac ctg ctg gac cag gtt ttc tgc			1119
Glu Glu Thr Tyr Val Lys Arg Leu His Leu Leu Asp Gln Val Phe Cys			
170	175	180	
acc agg ctg acg gat gcg ggg atc cct cca gaa gtc atc atg ggc ata			1167
Thr Arg Leu Thr Asp Ala Gly Ile Pro Pro Glu Val Ile Met Gly Ile			
185	190	195	200
ttc tct aac atc tcc tcc atc cac cgc ttc cac ggg cag ttc ctg ctg			1215
Phe Ser Asn Ile Ser Ser Ile His Arg Phe His Gly Gln Phe Leu Leu			
205	210	215	
ccg gag ctg aag acg cgg atc acg gag gag tgg gac aca aac cca cgg			1263
Pro Glu Leu Lys Thr Arg Ile Thr Glu Glu Trp Asp Thr Asn Pro Arg			
220	225	230	
ctc ggg gac atc ctg cag aag ctg gcc cca ttc ctg aag atg tac ggc			1311
Leu Gly Asp Ile Leu Gln Lys Leu Ala Pro Phe Leu Lys Met Tyr Gly			
235	240	245	
gag tat gtc aag aac ttt gac cga gcc gta ggg ctg gtg agc acg tgg			1359
Glu Tyr Val Lys Asn Phe Asp Arg Ala Val Gly Leu Val Ser Thr Trp			
250	255	260	
acc cag cgc tcc cca ctg ttt aaa gac gtc gtc cac agc atc cag aag			1407
Thr Gln Arg Ser Pro Leu Phe Lys Asp Val Val His Ser Ile Gln Lys			
265	270	275	280
cag gag gta tgc ggg aac ctg acg ctg cag cac cac atg ctg gag ccc			1455
Gln Glu Val Cys Gly Asn Leu Thr Leu Gln His His Met Leu Glu Pro			
285	290	295	
gtg cag agg gtc ccc cgg tac gag ctg ctg ctc aag gac tat ctg aag			1503
Val Gln Arg Val Pro Arg Tyr Glu Leu Leu Lys Asp Tyr Leu Lys			
300	305	310	
agg ctc ccg cag gac gcc cca gac cgg aag gat gcg gag agg tcc ttg			1551
Arg Leu Pro Gln Asp Ala Pro Asp Arg Lys Asp Ala Glu Arg Ser Leu			
315	320	325	
gag ctc atc tcc aca gcc gcc aac cac tcc aat gct gcc att cgg aaa			1599
Glu Leu Ile Ser Thr Ala Ala Asn His Ser Asn Ala Ala Ile Arg Lys			
330	335	340	
gtg gag aaa atg cac aag ctc ttg gag gtg tac gag cag ctg ggt ggg			1647
Val Glu Lys Met His Lys Leu Leu Glu Val Tyr Glu Gln Leu Gly Gly			
345	350	355	360
gaa gaa gac att gtc aac ccg gcc aat gaa ctg atc aag gag ggc caa			1695
Glu Glu Asp Ile Val Asn Pro Ala Asn Glu Leu Ile Lys Glu Gly Gln			
365	370	375	

atc cag aaa ctg tca gcc aag aac ggc acc ccc cag gac cgc cac ctc Ile Gln Lys Leu Ser Ala Lys Asn Gly Thr Pro Gln Asp Arg His Leu 380 385 390	1743
ttc ctg ttc aac agc atg atc ctt tac tgt gtg ccc aag ctg cgg ctc Phe Leu Phe Asn Ser Met Ile Leu Tyr Cys Val Pro Lys Leu Arg Leu 395 400 405	1791
atg ggc cag aag ttc agc gtc cgg gag aag atg gac atc tca ggc ctc Met Gly Gln Lys Phe Ser Val Arg Glu Lys Met Asp Ile Ser Gly Leu 410 415 420	1839
cag gtg cag gat atc gtc aag cca aac aca gca cat aca ttc atc ata Gln Val Gln Asp Ile Val Lys Pro Asn Thr Ala His Thr Phe Ile Ile 425 430 435 440	1887
acg gga aga aaa agg tcc ctg gag ctg cag acg cgg aca gag gaa gag Thr Gly Arg Lys Arg Ser Leu Glu Leu Gln Thr Arg Thr Glu Glu Glu 445 450 455	1935
aag aaa gaa tgg att cag atc atc cag gcc acc atc gag aag cac aaa Lys Lys Glu Trp Ile Gln Ile Ile Gln Ala Thr Ile Glu Lys His Lys 460 465 470	1983
cag aac agc gaa acc ttc aag gct ttt ggt ggc gcc ttc agc cag gat Gln Asn Ser Glu Thr Phe Lys Ala Phe Gly Gly Ala Phe Ser Gln Asp 475 480 485	2031
gag gac ccc agc ctc tct cca gac atg cct atc acg agc acc agc cct Glu Asp Pro Ser Leu Ser Pro Asp Met Pro Ile Thr Ser Thr Ser Pro 490 495 500	2079
gtg gag cct gtg gtg acc acc gaa ggc agt tcg ggt gca gca ggg ctc Val Glu Pro Val Val Thr Thr Glu Gly Ser Ser Gly Ala Ala Gly Leu 505 510 515 520	2127
gag ccc aga aaa cta tcc tct aag acc aga cgt gac aag gag aag cag Glu Pro Arg Lys Leu Ser Ser Lys Thr Arg Arg Asp Lys Glu Lys Gln 525 530 535	2175
agc tgt aag agc tgt ggt gag acc ttc aac tcc atc acc aag agg agg Ser Cys Lys Ser Cys Gly Glu Thr Phe Asn Ser Ile Thr Lys Arg Arg 540 545 550	2223
cat cac tgc aag ctg tgt ggg gcg gtc atc tgt ggg aag tgc tcc gag His His Cys Lys Leu Cys Gly Ala Val Ile Cys Gly Lys Cys Ser Glu 555 560 565	2271
ttc aag gcc gag aac agc cgg cag agc cgt gtc tgc aga gat tgt ttc Phe Lys Ala Glu Asn Ser Arg Gln Ser Arg Val Cys Arg Asp Cys Phe 570 575 580	2319
ctg aca cag cca gtg gcc cct gag agc aca gag aag aca ccc act gca Leu Thr Gln Pro Val Ala Pro Glu Ser Thr Glu Lys Thr Pro Thr Ala 585 590 595 600	2367

gac ccc cag ccc agc ctg ctc tgc ggc ccc ctg cgg ctg tca gag agc	2415
Asp Pro Gln Pro Ser Leu Leu Cys Gly Pro Leu Arg Leu Ser Glu Ser	
605 610 615	
gga gag acc tgg agc gag gtg tgg gcc gcc atc ccc atg tca gat ccc	2463
Gly Glu Thr Trp Ser Glu Val Trp Ala Ala Ile Pro Met Ser Asp Pro	
620 625 630	
cag gtg ctg cac ctg cag gga ggc agc cag gac ggc cgg ctg ccc cgc	2511
Gln Val Leu His Leu Gln Gly Gly Ser Gln Asp Gly Arg Leu Pro Arg	
635 640 645	
acc atc cct ctc ccc agc tgc aaa ctg agt gtg ccg gac cct gag gag	2559
Thr Ile Pro Leu Pro Ser Cys Lys Leu Ser Val Pro Asp Pro Glu Glu	
650 655 660	
agg ctg gac tcg ggg cat gtg tgg aag ctg cag tgg gcc aag cag tcc	2607
Arg Leu Asp Ser Gly His Val Trp Lys Leu Gln Trp Ala Lys Gln Ser	
665 670 675 680	
tgg tac ctg agc gcc tcc tcc gca gag ctg cag cag cag tgg ctg gaa	2655
Trp Tyr Leu Ser Ala Ser Ser Ala Glu Leu Gln Gln Gln Trp Leu Glu	
685 690 695	
acc cta agc act gct gcc cat ggg gac acg gcc cag gac agc ccg ggg	2703
Thr Leu Ser Thr Ala Ala His Gly Asp Thr Ala Gln Asp Ser Pro Gly	
700 705 710	
gcc ctg cag ctt cag gtc cct atg ggc gca gct gct ccg tga gctgagt	2752
Ala Leu Gln Leu Gln Val Pro Met Gly Ala Ala Ala Pro *	
715 720 725	
ctcccactgc cctgcacacc accacattgg acctgtgctg tcctgggagg tgggtgttga	2812
ggccccatga agagcgccct ggactgctga ggggtgggcca acagcccaga gctcaggaca	2872
tttggctttg gggggaagga aactgaggcc cagagagggg caaccactgg ccaaggggtca	2932
cccagcaagt tttggctaag agcctggcct ccagccccag cagtgtggcc cagagcaggg	2992
gccgactgcc aaagtaacca tcatccatat gggccgtgtg gtgatgctgg ccggaaggc	3052
agaaagaggc agcatgggca ctgccaggga cagccacatc ctgctgggtct gcagcgtggt	3112
ccaccccgcc cctgcccagc ctgtctacac cgtgtgagct gaatcgtgac ttgcttcca	3172
cctcctttct ctgtcctctc ctgaggttct gcctgcagcc cctaggaggt gggcctgccc	3232
catcctagct ggactcatgg ttctaaata accacgctca gaagctctgc taggacttac	3292
cccagccact gagtggcagg cgcattgaga	3321

<210> 7

<211> 2518

<212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> (332)..(2008)

<400> 7

```

aagctggtac gcctgcaggt accggtccgg aattcccggg tcgacttcgc tgtcgacgat      60
ttogtccgac tactgtactc ccaggcagaa acccacggct gaagcccacg atgcccattgt      120
gtaacttggt tattacacac acgtttggag cccagtgtgc agaccgcgag cgcagaggag      180
gcggcggaac accccggggt ggtcggggtt ccaagggctg acccgagctc cagcaactttt      240
tccgcgcctg atttttcaga gctcttcaga aaccaggctg ctttcaggaa cattgctgtg      300
gattcccagg gcctattcca ctagaagcaa g   atg gct gaa ctc aat act cat      352
                               Met Ala Glu Leu Asn Thr His
                               1               5

gtg aat gtc aag gaa aag atc tat gca gtt aga tca gtt gtt ccc aac      400
Val Asn Val Lys Glu Lys Ile Tyr Ala Val Arg Ser Val Val Pro Asn
      10               15               20

aaa agc aat aat gaa ata gtc ctg gtg ctc caa cag ttt gat ttt aat      448
Lys Ser Asn Asn Glu Ile Val Leu Val Leu Gln Gln Phe Asp Phe Asn
      25               30               35

gtg gat aaa gcc gtg caa gcc ttt gtg gat ggc agt gca att caa gtt      496
Val Asp Lys Ala Val Gln Ala Phe Val Asp Gly Ser Ala Ile Gln Val
      40               45               50               55

cta aaa gaa tgg aat atg aca gga aaa aag aag aac aat aaa aga aaa      544
Leu Lys Glu Trp Asn Met Thr Gly Lys Lys Lys Asn Asn Lys Arg Lys
      60               65               70

aga agc aag tcc aag cag cat caa ggc aac aaa gat gct aaa gac aag      592
Arg Ser Lys Ser Lys Gln His Gln Gly Asn Lys Asp Ala Lys Asp Lys
      75               80               85

gtg gag agg cct gag gca ggg ccc ctg cag ccg cag cca cca cag att      640
Val Glu Arg Pro Glu Ala Gly Pro Leu Gln Pro Gln Pro Pro Gln Ile
      90               95               100

caa aac ggc ccc atg aat ggc tgc gag aag gac agc tcg tcc aca gat      688
Gln Asn Gly Pro Met Asn Gly Cys Glu Lys Asp Ser Ser Ser Thr Asp
      105               110               115

tct gct aac gaa aaa cca gcc ctt atc cct cgt gag aaa aag atc tcg      736
Ser Ala Asn Glu Lys Pro Ala Leu Ile Pro Arg Glu Lys Lys Ile Ser
      120               125               130               135

ata ctt gag gaa cct tca aag gca ctt cgt ggg gtc aca gaa ggc aac      784
Ile Leu Glu Glu Pro Ser Lys Ala Leu Arg Gly Val Thr Glu Gly Asn
      140               145               150

```

aga cta ctg caa cag aaa cta tcc tta gat ggg aac ccc aaa cct ata	832
Arg Leu Leu Gln Gln Lys Leu Ser Leu Asp Gly Asn Pro Lys Pro Ile	
155 160 165	
cat gga aca aca gag agg tca gat ggc cta cag tgg tca gct gag cag	880
His Gly Thr Thr Glu Arg Ser Asp Gly Leu Gln Trp Ser Ala Glu Gln	
170 175 180	
cct tgt aac cca agc aag cct aag gca aaa aca tct cct gtt aag tcc	928
Pro Cys Asn Pro Ser Lys Pro Lys Ala Lys Thr Ser Pro Val Lys Ser	
185 190 195	
aat acc cct gca gct cat ctt gaa ata aag cca gat gag ttg gca aag	976
Asn Thr Pro Ala Ala His Leu Glu Ile Lys Pro Asp Glu Leu Ala Lys	
200 205 210 215	
aaa aga ggc cca aat att gag aaa tca gtg aag gat ttg caa cgc tgc	1024
Lys Arg Gly Pro Asn Ile Glu Lys Ser Val Lys Asp Leu Gln Arg Cys	
220 225 230	
acc gtt tct cta act aga tat cgc gtc atg att aag gaa gaa gtg gat	1072
Thr Val Ser Leu Thr Arg Tyr Arg Val Met Ile Lys Glu Glu Val Asp	
235 240 245	
agt tcc gtg aag aag atc aaa gct gcc ttt gct gaa tta cac aac tgc	1120
Ser Ser Val Lys Lys Ile Lys Ala Ala Phe Ala Glu Leu His Asn Cys	
250 255 260	
atc att gac aaa gaa gtt tca tta atg gca gaa atg gat aaa gtt aaa	1168
Ile Ile Asp Lys Glu Val Ser Leu Met Ala Glu Met Asp Lys Val Lys	
265 270 275	
gaa gaa gcc atg gaa atc ctg act gct cgt cag aag aaa gca gaa gaa	1216
Glu Glu Ala Met Glu Ile Leu Thr Ala Arg Gln Lys Lys Ala Glu Glu	
280 285 290 295	
cta aag aga ctc act gac ctt gcc agt cag atg gca gag atg cag ctg	1264
Leu Lys Arg Leu Thr Asp Leu Ala Ser Gln Met Ala Glu Met Gln Leu	
300 305 310	
gcc gaa ctc agg gca gaa att aag cac ttt gtc agc gag cgt aaa tat	1312
Ala Glu Leu Arg Ala Glu Ile Lys His Phe Val Ser Glu Arg Lys Tyr	
315 320 325	
gac gag gag ctc ggg aaa gct gcc cgg ttt tcc tgt gac atc gaa cag	1360
Asp Glu Glu Leu Gly Lys Ala Ala Arg Phe Ser Cys Asp Ile Glu Gln	
330 335 340	
ctg aag gcc caa atc atg ctc tgc gga gaa att aca cat cca aag aac	1408
Leu Lys Ala Gln Ile Met Leu Cys Gly Glu Ile Thr His Pro Lys Asn	
345 350 355	
aac tat tcc tca aga act ccc tgc agc tcc ctg ctg cct ctg ctg aat	1456
Asn Tyr Ser Ser Arg Thr Pro Cys Ser Ser Leu Leu Pro Leu Leu Asn	
360 365 370 375	

gcg	cac	gca	acc	tct	ggg	aaa	cag	agt	aac	ttt	tcc	cga	aaa	tca			
Ala	His	Ala	Ala	Thr	Ser	Gly	Lys	Gln	Ser	Asn	Phe	Ser	Arg	Lys	Ser		
380								385				390					
tcc	act	cac	aat	aag	ccc	tct	gaa	ggc	aaa	gcg	gca	aac	ccc	aaa	atg		
Ser	Thr	His	Asn	Lys	Pro	Ser	Glu	Gly	Lys	Ala	Ala	Asn	Pro	Lys	Met		
395								400				405					
gtg	agc	agt	ctc	ccc	agc	acc	gcc	gac	ccc	tct	cac	cag	acc	atg	cCG		
Val	Ser	Ser	Leu	Pro	Ser	Thr	Ala	Asp	Pro	Ser	His	Gln	Thr	Met	Pro		
410								415				420					
gcc	aac	aag	cag	aat	gga	tct	tct	aac	caa	aga	cgg	aga	ttt	aat	cca		
Ala	Asn	Lys	Gln	Asn	Gly	Ser	Ser	Asn	Gln	Arg	Arg	Arg	Phe	Asn	Pro		
425								430				435					
cag	tat	cat	aac	aac	agg	cta	aat	ggg	cct	gcc	aag	tcg	cag	ggc	agt		
Gln	Tyr	His	Asn	Asn	Arg	Leu	Asn	Gly	Pro	Ala	Lys	Ser	Gln	Gly	Ser		
440								445				450					
ggg	aat	gaa	gcc	gag	cca	ctg	gga	aag	ggc	aac	agc	cgc	cac	gaa	cac		
Gly	Asn	Glu	Ala	Glu	Pro	Leu	Gly	Lys	Gly	Asn	Ser	Arg	His	Glu	His		
460								465				470					
aga	aga	cag	ccg	cac	aac	ggc	ttc	cgg	ccc	aaa	aac	aaa	ggc	ggg	gcc		
Arg	Arg	Gln	Pro	His	Asn	Gly	Phe	Arg	Pro	Lys	Asn	Lys	Gly	Gly	Ala		
475								480				485					
aaa	aat	caa	gag	gct	tcc	ttg	ggg	atg	aag	acc	ccc	gag	gcc	ccg	gcc		
Lys	Asn	Gln	Glu	Ala	Ser	Leu	Gly	Met	Lys	Thr	Pro	Glu	Ala	Pro	Ala		
490								495				500					
cat	tct	gaa	aag	ccc	cgg	cga	agg	cag	cac	gct	gca	gac	acc	tcg	gag		
His	Ser	Glu	Lys	Pro	Arg	Arg	Arg	Gln	His	Ala	Ala	Asp	Thr	Ser	Glu		
505								510				515					
gcc	agg	ccc	ttc	cgg	ggt	agt	gtc	ggt	agg	gtt	tca	cag	tgc	aat	ctc		
Ala	Arg	Pro	Phe	Arg	Gly	Ser	Val	Gly	Arg	Val	Ser	Gln	Cys	Asn	Leu		
520								525				530					
tgc	ccc	acg	aga	ata	gaa	gtt	tcc	aca	gat	gca	gca	gtt	ctc	tca	gtc		
Cys	Pro	Thr	Arg	Ile	Glu	Val	Ser	Thr	Asp	Ala	Ala	Val	Leu	Ser	Val		
540								545				550					
ccg	gct	gtg	acg	ttg	gtg	gcc	tga	gctaggagga				aaaagagcag				ttttcactca	
Pro	Ala	Val	Thr	Leu	Val	Ala	*										
555																	
gttttggttc cctgcccgag gtgctgacct aattcgctgc caaaagagtg tcaatcagaa																	
2098																	
tatacaaatc ccgtatggtt gtgtcatcct ctcttaatca tttttactaa ttctaataat																	
2158																	
cagctctagc ttgcttcata attttcatgg ctttgcttga tctgttgatg ctttctctca																	
2218																	
tcaagacttt gcagcatttt agccaggcag tattttactca ttattaggaa aatcaagatg																	
2278																	
tggetgaaga tcagaggctc agtttagaac ctgtgttgta gcagtqatgt cagtccattg																	
2338																	

attgtcttta gagagttaat gttacaaaaa agaattctta ataatcagac aaacatgatc 2398  
 tgctgaggac acatgcgctt ttgtagaatt taacatctgg tgtttttctg aaaaaatata 2458  
 tatacatata ttgctttatt tgaacaaat taaaatatgc tgcatttgaa aaaaaaaaaa 2518

<210> 8  
 <211> 2311  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> (332)..(1801)

<400> 8  
 aagctggtac gcctgcaggt accggtccgg aattccccggg tcgacttcgc tgcgcacgat 60  
 ttcgctccgac tactgtactc ccaggcagaa acccacggct gaagcccacg atgcccattgt 120  
 gtaacttggg tattacacac acgtttggag cccagtgtgc agaccgcag cgcagaggag 180  
 gcggcggaac accccggggt ggctcgggttt ccaagggctg acccgagctc cagcactttt 240  
 tccgcgcctg atttttcaga gctcttcaga aaccaggctg ctttcaggaa cattgctgtg 300  
 gattcccagg gcctattcca ctagaagcaa g atg gct gaa ctc aat act cat 352  
 Met Ala Glu Leu Asn Thr His  
 1 5  
 gtg aat gtc aag gaa aag atc tat gca gtt aga tca gtt gtt ccc aac 400  
 Val Asn Val Lys Glu Lys Ile Tyr Ala Val Arg Ser Val Val Pro Asn  
 10 15 20  
 aaa agc aat aat gaa ata gtc ctg gtg ctc caa cag ttt gat ttt aat 448  
 Lys Ser Asn Asn Glu Ile Val Leu Val Leu Gln Phe Asp Phe Asn  
 25 30 35  
 gtg gat aaa gcc gtg caa gcc ttt gtg gat ggc agt gca att caa gtt 496  
 Val Asp Lys Ala Val Gln Ala Phe Val Asp Gly Ser Ala Ile Gln Val  
 40 45 50 55  
 cta aaa gaa tgg aat atg aca gga aaa aag aag aac aat aaa aga aaa 544  
 Leu Lys Glu Trp Asn Met Thr Gly Lys Lys Lys Asn Asn Lys Arg Lys  
 60 65 70  
 aga agc aag tcc aag cag cat caa ggc aac aaa gat gct aaa gac aag 592  
 Arg Ser Lys Ser Lys Gln His Gln Gly Asn Lys Asp Ala Lys Asp Lys  
 75 80 85  
 gtg gag agg cct gag gca ggg ccc ctg cag ccg cag cca cca cag att 640  
 Val Glu Arg Pro Glu Ala Gly Pro Leu Gln Pro Gln Pro Pro Gln Ile  
 90 95 100



caa Gln	aac Asn	ggc Gly	ccc Pro	atg Met	aat Asn	ggc Gly	tgc Cys	gag Glu	aag Lys	gac Asp	agc Ser	tcg Ser	tcc Ser	aca Thr	gat Asp	688
105110115																
tct Ser	gct Ala	aac Asn	gaa Glu	aaa Lys	cca Pro	gcc Ala	ctt Leu	atc Ile	cct Pro	cgt Arg	gag Glu	aaa Lys	aag Lys	atc Ile	tcg Ser	736
120125130135																
ata Ile	ctt Leu	gag Glu	gaa Glu	cct Pro	tca Ser	aag Lys	gca Ala	ctt Leu	cgt Arg	ggg Gly	gtc Val	aca Thr	ggc Gly	cca Pro	aat Asn	784
140145150																
att Ile	gag Glu	aaa Lys	tca Ser	gtg Val	aag Lys	gat Asp	ttg Leu	caa Gln	cgc Arg	tgc Cys	acc Thr	gtt Val	tct Ser	cta Leu	act Thr	832
155160165																
aga Arg	tat Tyr	cgc Arg	gtc Val	atg Met	att Ile	aag Lys	gaa Glu	gaa Glu	gtg Val	gat Asp	agt Ser	tcc Ser	gtg Val	aag Lys	aag Lys	880
170175180																
atc Ile	aaa Lys	gct Ala	gcc Ala	ttt Phe	gct Ala	gaa Glu	tta Leu	cac His	aac Asn	tgc Cys	atc Ile	att Ile	gac Asp	aaa Lys	gaa Glu	928
185190195																
gtt Val	tca Ser	tta Leu	atg Met	gca Ala	gaa Glu	atg Met	gat Asp	aaa Lys	gtt Val	aaa Lys	gaa Glu	gaa Glu	gcc Ala	atg Met	gaa Glu	976
200205210215																
atc Ile	ctg Leu	act Thr	gct Ala	cgt Arg	cag Gln	aag Lys	aaa Lys	gca Ala	gaa Glu	gaa Glu	cta Leu	aag Lys	aga Arg	ctc Leu	act Thr	1024
220225230																
gac Asp	ctt Leu	gcc Ala	agt Ser	cag Gln	atg Met	gca Ala	gag Glu	atg Met	cag Gln	ctg Leu	gcc Ala	gaa Glu	ctc Leu	agg Arg	gca Ala	1072
235240245																
gaa Glu	att Ile	aag Lys	cac His	ttt Phe	gtc Val	agc Ser	gag Glu	cgt Arg	aaa Lys	tat Tyr	gac Asp	gag Glu	gag Glu	ctc Leu	ggg Gly	1120
250255260																
aaa Lys	gct Ala	gcc Ala	cgg Arg	ttt Phe	tcc Ser	tgt Cys	gac Asp	atc Ile	gaa Glu	cag Gln	ctg Leu	aag Lys	gcc Ala	caa Gln	atc Ile	1168
265270275																
atg Met	ctc Leu	tgc Cys	gga Gly	gaa Glu	att Ile	aca Thr	cat His	cca Pro	aag Lys	aac Asn	aac Asn	tat Tyr	tcc Ser	tca Ser	aga Arg	1216
280285290295																
act Thr	ccc Pro	tgc Cys	agc Ser	tcc Ser	ctg Leu	ctg Leu	cct Pro	ctg Leu	ctg Leu	aat Asn	gcg Ala	cac His	gca Ala	gca Ala	acc Thr	1264
300305310																
tct Ser	ggg Gly	aaa Lys	cag Gln	agt Ser	aac Asn	ttt Phe	tcc Ser	cga Arg	aaa Lys	tca Ser	tcc Ser	act Thr	cac His	aat Asn	aag Lys	1312
315320325																

ccc tct gaa ggc aaa gcg gca aac ccc aaa atg gtg agc agt ctc ccc Pro Ser Glu Gly Lys Ala Ala Asn Pro Lys Met Val Ser Ser Leu Pro 330 335 340	1360
agc acc gcc gac ccc tct cac cag acc atg ccg gcc aac aag cag aat Ser Thr Ala Asp Pro Ser His Gln Thr Met Pro Ala Asn Lys Gln Asn 345 350 355	1408
gga tct tct aac caa aga cgg aga ttt aat cca cag tat cat aac aac Gly Ser Ser Asn Gln Arg Arg Arg Phe Asn Pro Gln Tyr His Asn Asn 360 365 370 375	1456
agg cta aat ggg cct gcc aag tcg cag ggc agt ggg aat gaa gcc gag Arg Leu Asn Gly Pro Ala Lys Ser Gln Gly Ser Gly Asn Glu Ala Glu 380 385 390	1504
cca ctg gga aag ggc aac agc cgc cac gaa cac aga aga cag ccg cac Pro Leu Gly Lys Gly Asn Ser Arg His Glu His Arg Arg Gln Pro His 395 400 405	1552
aac ggc ttc cgg ccc aaa aac aaa ggc ggt gcc aaa aat caa gag gct Asn Gly Phe Arg Pro Lys Asn Lys Gly Gly Ala Lys Asn Gln Glu Ala 410 415 420	1600
tcc ttg ggg atg aag acc ccc gag gcc ccg gcc cat tct gaa aag ccc Ser Leu Gly Met Lys Thr Pro Glu Ala Pro Ala His Ser Glu Lys Pro 425 430 435	1648
cgg cga agg cag cac gct gca gac acc tcg gag gcc agg ccc ttc cgg Arg Arg Arg Gln His Ala Ala Asp Thr Ser Glu Ala Arg Pro Phe Arg 440 445 450 455	1696
ggt agt gtc ggt agg gtt tca cag tgc aat ctc tgc ccc acg aga ata Gly Ser Val Gly Arg Val Ser Gln Cys Asn Leu Cys Pro Thr Arg Ile 460 465 470	1744
gaa gtt tcc aca gat gca gca gtt ctc tca gtc ccg gct gtg acg ttg Glu Val Ser Thr Asp Ala Ala Val Leu Ser Val Pro Ala Val Thr Leu 475 480 485	1792
gtg gcc tga gctagga ggaaaaagag cagttttcac tcagttttgg ttccctgccc Val Ala * 490	1848
gaggtgctga cccaattcgc tgccaaaaga gtgtcaatca gaatatacaa atcccgtatg	1908
gttgtgtcat cctctcttaa tcattttttac taattctaat aatcagctct agcttgcttc	1968
ataattttca tggctttgct tgatctgttg atgctttctc tcatcaagac tttgcagcat	2028
tttagccagg cagtattttac tcattatttag gaaaatcaag atgtggctga agatcagagg	2088
ctcagttagc aacctgtgtt gtagcagtga tgtcagtcca ttgattgtct ttagagagtt	2148
aatgttaciaa aaaagaattc ttaataatca gacaaacatg atctgctgag gacacatgcg	2208
cttttgtaga atttaacatc tgggtgttttt ctgaaaaaat atatatacat atattgcttt	2268

atttgaaaca aattaaaata tgctgcattt gaaaaaaaaa aaa 2311

<210> 9  
<211> 746  
<212> DNA  
<213> Homo sapiens

<220>  
<221> CDS  
<222> (280)..(627)

<400> 9

atttggccct cgaggccaag aattcggcac gagcttgaag ggaggtagta accaagggttc	60
cgctggctcct gagacttctt ggggcctgat ttgattaacc caggaagagt ctcagacctc	120
atgaagtact gccaggtctg gcagcacttg gccaaagcgg gcaggtctgt gttgaggttg	180
ctttatgcag actaaciaag ggatcattct gctgtgtacc tgtccctatt ctagacctg	240
gtgggcagca gtgggacct tctgaccaca atgcctgta atg atg ggg caa gag	294
Met Met Gly Gln Glu	
1 5	
aaa gtg ccc att aag cag gta cct ggg gga gtc aag cag ctt gag ccc	342
Lys Val Pro Ile Lys Gln Val Pro Gly Gly Val Lys Gln Leu Glu Pro	
10 15 20	
ccc aaa gaa gga gaa agg cgg aca acc cat aat atc att gag aaa cga	390
Pro Lys Glu Gly Glu Arg Arg Thr Thr His Asn Ile Ile Glu Lys Arg	
25 30 35	
tat cgc tcc tcc atc aat gac aaa atc atc gaa ttg aaa gac ctg gtc	438
Tyr Arg Ser Ser Ile Asn Asp Lys Ile Ile Glu Leu Lys Asp Leu Val	
40 45 50	
atg ggg aca gac gcc aag atg cac aag tct ggc gtt ctg agg aag gcc	486
Met Gly Thr Asp Ala Lys Met His Lys Ser Gly Val Leu Arg Lys Ala	
55 60 65	
att gat tac acc aga tgc ctg aaa ggc ctc cat ggc ttt cct gcc ctt	534
Ile Asp Tyr Thr Arg Cys Leu Lys Gly Leu His Gly Phe Pro Ala Leu	
70 75 80 85	
cct ggt tcc gga cag ctg ggg aaa ggc cac agc agc tcc tct gct gcc	582
Pro Gly Ser Gly Gln Leu Gly Lys Gly His Ser Ser Ser Ser Ala Ala	
90 95 100	
ctg cag tct ttg ggg gcg ggg agg gct gga cat gtg gaa ccc tga tgc	630
Leu Gln Ser Leu Gly Ala Gly Arg Ala Gly His Val Glu Pro *	
105 110 115	
agccgcagcg tcaaggacga ggaaggggtg ggaagggatg gtacgtggag gggaatgggt	690

gggtgggacca gggacccaat gctaataaag actggactgt gcttcaaaaa aaaaaa 746

<210> 10  
<211> 3580  
<212> DNA  
<213> Homo sapiens

<220>  
<221> CDS  
<222> (373)..(3108)

<220>  
<221> misc\_feature  
<222> (1)...(3580)  
<223> n = a,t,c or g

<400> 10  
catagacagt cgatacgctt gcacgnaccg gnccggaatt cccgggtcga cgatttcgtc 60  
gagtgaaga cgggcgggct gctgagggcg ccccgcggtt ctggtctggg ccgggggaact 120  
ccgggacggc gccgcagcag ggctacggat ctctcggccc ctgcccttca aatcacctct 180  
aggcacaggt tattcagctt ctcaaggctg gaaaagcgaa ggaagtttcc tacaatgcac 240  
tagcctcaca cataatctca gaggatgggg acaatccaga ggtgggagaa gctcgggaag 300  
tctttgactt acctgttgta aagccttctt ggggtgattct gtccgttcag tgtggaactc 360  
ttctgccagt aa atg gtt ttt ctc cag aat cat gtc aga ttt ttt ttg 408  
Met Val Phe Leu Gln Asn His Val Arg Phe Phe Leu  
1 5 10  
gaa tca ctg cct gcc ttt ctc agg gtg ttg ata caa gct gga gct ctt 456  
Glu Ser Leu Pro Ala Phe Leu Arg Val Leu Ile Gln Ala Gly Ala Leu  
15 20 25  
tgt tgg agt ctt cca gag ctc tcc caa gga gag gta ggg aag gga gct 504  
Cys Trp Ser Leu Pro Glu Leu Ser Gln Gly Glu Val Gly Lys Gly Ala  
30 35 40  
tgt cca gca gaa gtt ggg aag cac aga gat cat ctg cct tct tct gac 552  
Cys Pro Ala Glu Val Gly Lys His Arg Asp His Leu Pro Ser Ser Asp  
45 50 55 60  
ccg gta ttg atg cag gct gag gcc tct gtt gta atg tgc tgg gtg tca 600  
Pro Val Leu Met Gln Ala Glu Ala Ser Val Val Met Cys Trp Val Ser  
65 70 75  
tct gaa gac aga agt gcc ctg tgg gct ttg gtt acg ttc tat ggg gga 648  
Ser Glu Asp Arg Ser Ala Leu Trp Ala Leu Val Thr Phe Tyr Gly Gly  
80 85 90

gat tgc cag cta acc ctc aat aag aaa tgc acg cat ttg att gtt cca	696
Asp Cys Gln Leu Thr Leu Asn Lys Lys Cys Thr His Leu Ile Val Pro	
95 100 105	
gag cca aag ggg ggg aaa tac gaa tgt gct tta aag cga gca agt att	744
Glu Pro Lys Gly Gly Lys Tyr Glu Cys Ala Leu Lys Arg Ala Ser Ile	
110 115 120	
aaa att gtg act cct gac tgg gtt ctg gat tgc gta tca gag aaa acc	792
Lys Ile Val Thr Pro Asp Trp Val Leu Asp Cys Val Ser Glu Lys Thr	
125 130 135 140	
aaa aag gac gaa gca ttt tat cat cct cgt ctg att att tat gaa gag	840
Lys Lys Asp Glu Ala Phe Tyr His Pro Arg Leu Ile Ile Tyr Glu Glu	
145 150 155	
gaa gaa gag gaa gag gaa gag gag gag gaa gta gaa aat gag gaa caa	888
Glu Glu Glu Glu Glu Glu Glu Glu Glu Glu Val Glu Asn Glu Glu Gln	
160 165 170	
gat tct cag aat gag ggt agt aca gat gag aag tca agc cct gcc agc	936
Asp Ser Gln Asn Glu Gly Ser Thr Asp Glu Lys Ser Ser Pro Ala Ser	
175 180 185	
tct caa gaa ggg tct cct tca ggt gac cag cag ttt tca cct aaa tcc	984
Ser Gln Glu Gly Ser Pro Ser Gly Asp Gln Gln Phe Ser Pro Lys Ser	
190 195 200	
aac act gaa aaa tct aaa ggg gaa tta atg ttt gat gat tct tca gat	1032
Asn Thr Glu Lys Ser Lys Gly Glu Leu Met Phe Asp Asp Ser Ser Asp	
205 210 215 220	
tca tca ccg gaa aaa cag gag aga aat tta aac tgg acc ccg gcc gaa	1080
Ser Ser Pro Glu Lys Gln Glu Arg Asn Leu Asn Trp Thr Pro Ala Glu	
225 230 235	
gtc cca cag tta gct gca gca aaa cgc agg ctg cct cag gga aag gag	1128
Val Pro Gln Leu Ala Ala Ala Lys Arg Arg Leu Pro Gln Gly Lys Glu	
240 245 250	
cct ggg ttg att aac ttg tgt gcc aat gtc cca ccc gtc cca ggt aac	1176
Pro Gly Leu Ile Asn Leu Cys Ala Asn Val Pro Pro Val Pro Gly Asn	
255 260 265	
att ttg ccc cct gag gtc cgg ggt aat tta atg gct gct gga caa aac	1224
Ile Leu Pro Pro Glu Val Arg Gly Asn Leu Met Ala Ala Gly Gln Asn	
270 275 280	
ctc caa agt tct gaa aga tca gaa atg ata gct acc tgg agt cca gct	1272
Leu Gln Ser Ser Glu Arg Ser Glu Met Ile Ala Thr Trp Ser Pro Ala	
285 290 295 300	
gta cgg aca ctg agg aat att act aat aat gct gac att cag cag atg	1320
Val Arg Thr Leu Arg Asn Ile Thr Asn Asn Ala Asp Ile Gln Gln Met	
305 310 315	
aac cgg cca tca aat gta gca cat atc tta cag act ctt tca gca cct	1368

Asn	Arg	Pro	Ser	Asn	Val	Ala	His	Ile	Leu	Gln	Thr	Leu	Ser	Ala	Pro	
			320					325					330			
acg	aaa	aat	tta	gaa	cag	cag	gtg	aat	cac	agc	cag	cag	gga	cat	aca	1416
Thr	Lys	Asn	Leu	Glu	Gln	Gln	Val	Asn	His	Ser	Gln	Gln	Gly	His	Thr	
		335					340					345				
aat	gcc	aat	gca	gtg	ctg	ttt	agc	caa	gtg	aaa	gtg	act	cca	gag	aca	1464
Asn	Ala	Asn	Ala	Val	Leu	Phe	Ser	Gln	Val	Lys	Val	Thr	Pro	Glu	Thr	
	350					355					360					
cac	atg	cta	cag	cag	cag	cag	cag	gcc	cag	cag	cag	cag	cag	cag	cac	1512
His	Met	Leu	Gln	Gln	Gln	Gln	Gln	Ala	Gln	Gln	Gln	Gln	Gln	Gln	His	
365					370					375					380	
ccg	gtt	tta	cac	ctt	cag	ccc	cag	cag	ata	atg	cag	ctc	cag	cag	cag	1560
Pro	Val	Leu	His	Leu	Gln	Pro	Gln	Gln	Ile	Met	Gln	Leu	Gln	Gln	Gln	
				385					390					395		
cag	cag	cag	cag	atc	tct	cag	caa	cct	tac	ccc	cag	cag	ccg	ccg	cat	1608
Gln	Gln	Gln	Gln	Ile	Ser	Gln	Gln	Pro	Tyr	Pro	Gln	Gln	Pro	Pro	His	
			400					405					410			
cca	ttt	tca	cag	cag	cag	cag	cag	caa	cag	cag	cca	cca	cca	tcg	cct	1656
Pro	Phe	Ser	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Pro	Pro	Pro	Ser	Pro	
		415					420					425				
cag	cag	cat	cag	ctt	ttt	gga	cat	gat	cca	gca	gtg	gag	att	cca	gaa	1704
Gln	Gln	His	Gln	Leu	Phe	Gly	His	Asp	Pro	Ala	Val	Glu	Ile	Pro	Glu	
	430					435					440					
gaa	ggc	ttc	tta	ttg	gga	tgt	gtg	ttt	gca	att	gcg	gat	tat	cca	gag	1752
Glu	Gly	Phe	Leu	Leu	Gly	Cys	Val	Phe	Ala	Ile	Ala	Asp	Tyr	Pro	Glu	
445				450					455					460		
cag	atg	tct	gat	aag	caa	ctg	ctg	gcc	acc	tgg	aaa	agg	ata	atc	cag	1800
Gln	Met	Ser	Asp	Lys	Gln	Leu	Leu	Ala	Thr	Trp	Lys	Arg	Ile	Ile	Gln	
			465					470					475			
gca	cat	ggc	ggc	act	gtt	gac	ccc	acc	ttc	acg	agt	cga	tgc	acg	cac	1848
Ala	His	Gly	Gly	Thr	Val	Asp	Pro	Thr	Phe	Thr	Ser	Arg	Cys	Thr	His	
		480					485					490				
ctt	ctc	tgt	gag	agt	caa	gtc	agc	agc	gcg	tat	gca	cag	gca	ata	aga	1896
Leu	Leu	Cys	Glu	Ser	Gln	Val	Ser	Ser	Ala	Tyr	Ala	Gln	Ala	Ile	Arg	
		495				500					505					
gaa	aga	aag	aga	tgt	gtt	act	gca	cac	tgg	tta	aac	aca	gtc	tta	aaa	1944
Glu	Arg	Lys	Arg	Cys	Val	Thr	Ala	His	Trp	Leu	Asn	Thr	Val	Leu	Lys	
	510					515					520					
aaa	aaa	aaa	atg	gta	ccg	ccg	cac	cga	gcc	ctt	cac	ttc	cca	gtg	gcc	1992
Lys	Lys	Lys	Met	Val	Pro	Pro	His	Arg	Ala	Leu	His	Phe	Pro	Val	Ala	

545																550																555																
gga	ttt	gtt	gat	agt	gac	aga	gat	gac	cta	aaa	tta	atg	gct	tat	ttg	2088																																
Gly	Phe	Val	Asp	Ser	Asp	Arg	Asp	Asp	Leu	Lys	Leu	Met	Ala	Tyr	Leu																																	
			560						565			570																																				
gca	ggt	gcc	aaa	tat	acg	ggt	tat	cta	tgc	cgc	agc	aac	aca	gtc	ctc	2136																																
Ala	Gly	Ala	Lys	Tyr	Thr	Gly	Tyr	Leu	Cys	Arg	Ser	Asn	Thr	Val	Leu																																	
			575			580						585																																				
atc	tgt	aaa	gaa	cca	act	ggt	tta	aag	tat	gaa	aaa	gcc	aaa	gag	tgg	2184																																
Ile	Cys	Lys	Glu	Pro	Thr	Gly	Leu	Lys	Tyr	Glu	Lys	Ala	Lys	Glu	Trp																																	
			590			595						600																																				
agg	ata	ccc	tgt	gtc	aac	gcc	cag	tgg	ctt	ggc	gac	att	ctt	ctg	gga	2232																																
Arg	Ile	Pro	Cys	Val	Asn	Ala	Gln	Trp	Leu	Gly	Asp	Ile	Leu	Leu	Gly																																	
605						610						615			620																																	
aac	ttt	gag	gca	ctg	agg	cag	att	cag	tat	agt	cgc	tac	acg	gca	ttc	2280																																
Asn	Phe	Glu	Ala	Leu	Arg	Gln	Ile	Gln	Tyr	Ser	Arg	Tyr	Thr	Ala	Phe																																	
			625						630						635																																	
agt	ctg	cag	gat	cca	ttt	gcc	cct	acc	cag	cat	tta	gtt	tta	aat	ctt	2328																																
Ser	Leu	Gln	Asp	Pro	Phe	Ala	Pro	Thr	Gln	His	Leu	Val	Leu	Asn	Leu																																	
			640						645						650																																	
tta	gat	gct	tgg	aga	gtt	ccc	tta	aaa	gtg	tct	gca	gag	ttg	ttg	atg	2376																																
Leu	Asp	Ala	Trp	Arg	Val	Pro	Leu	Lys	Val	Ser	Ala	Glu	Leu	Leu	Met																																	
			655			660						665																																				
agt	ata	aga	cta	cct	ccc	aaa	ctg	aaa	cag	aat	gaa	gta	gct	aat	gtc	2424																																
Ser	Ile	Arg	Leu	Pro	Pro	Lys	Leu	Lys	Gln	Asn	Glu	Val	Ala	Asn	Val																																	
			670			675						680																																				
cag	cct	tct	tcc	caa	aga	gcc	aga	att	gaa	gac	gta	cca	cct	ccc	act	2472																																
Gln	Pro	Ser	Ser	Gln	Arg	Ala	Arg	Ile	Glu	Asp	Val	Pro	Pro	Pro	Thr																																	
685						690						695			700																																	
aaa	aag	cta	act	cca	gaa	ttg	acc	cct	ttt	gtg	ctt	ttc	act	gga	ttc	2520																																
Lys	Lys	Leu	Thr	Pro	Glu	Leu	Thr	Pro	Phe	Val	Leu	Phe	Thr	Gly	Phe																																	
			705						710						715																																	
gag	cct	gtc	cag	gtt	caa	cag	tat	att	aag	aag	ctc	tac	att	ctt	ggt	2568																																
Glu	Pro	Val	Gln	Val	Gln	Gln	Tyr	Ile	Lys	Lys	Leu	Tyr	Ile	Leu	Gly																																	
			720						725						730																																	
gga	gag	gtt	gcg	gag	tct	gca	cag	aag	tgc	aca	cac	ctc	att	gcc	agc	2616																																
Gly	Glu	Val	Ala	Glu	Ser	Ala	Gln	Lys	Cys	Thr	His	Leu	Ile	Ala	Ser																																	
			735			740						745																																				
aaa	gtg	act	cgc	acc	gtg	aag	ttc	ctg	acg	gcg	att	tct	gtc	gtg	aag	2664																																
Lys	Val	Thr	Arg	Thr	Val	Lys	Phe	Leu	Thr	Ala	Ile	Ser	Val	Val	Lys																																	
			750			755						760																																				
cac	ata	gtg	acg	cca	gag	tgg	ctg	gaa	gaa	tgc	ttc	agg	tgt	cag	aag	2712																																
His	Ile	Val	Thr	Pro	Glu	Trp	Leu	Glu	Glu	Cys	Phe	Arg	Cys	Gln	Lys																																	
765						770						775			780																																	

ttc att gat gag cag aac tac att ctc cga gat gct gag gca gaa gta	2760
Phe Ile Asp Glu Gln Asn Tyr Ile Leu Arg Asp Ala Glu Ala Glu Val	
785 790 795	
ctt ttc tct ttc agc ttg gaa gaa tcc tta aaa cgg gca cac gtt tct	2808
Leu Phe Ser Phe Ser Leu Glu Glu Ser Leu Lys Arg Ala His Val Ser	
800 805 810	
cca ctc ttt aag gca aaa tat ttt tac atc aca cct gga atc tgc cca	2856
Pro Leu Phe Lys Ala Lys Tyr Phe Tyr Ile Thr Pro Gly Ile Cys Pro	
815 820 825	
agt ctt tcc act atg aag gca atc gta gag tgt gca gga gga aag gtg	2904
Ser Leu Ser Thr Met Lys Ala Ile Val Glu Cys Ala Gly Gly Lys Val	
830 835 840	
tta tcc aag cag cca tct ttc cgg aag ctc atg gag cac aag cag aac	2952
Leu Ser Lys Gln Pro Ser Phe Arg Lys Leu Met Glu His Lys Gln Asn	
845 850 855 860	
tcg agt ttg tcg gaa ata att tta ata tcc tgt gaa aat gac ctt cat	3000
Ser Ser Leu Ser Glu Ile Ile Leu Ile Ser Cys Glu Asn Asp Leu His	
865 870 875	
tta tgc cga gaa tat ttt gcc aga ggc ata gat gtt cac aat gca gag	3048
Leu Cys Arg Glu Tyr Phe Ala Arg Gly Ile Asp Val His Asn Ala Glu	
880 885 890	
ttc gtt ctg act gga gtg ctc act caa acg ctg gac tat gaa tca tat	3096
Phe Val Leu Thr Gly Val Leu Thr Gln Thr Leu Asp Tyr Glu Ser Tyr	
895 900 905	
aag ttt aac tga tgg cgtctaggct gccgtgcatg tcgactcctg cgggtgcgggg	3151
Lys Phe Asn *	
910	
ctggctgtct ggctggcgag gagctgctgc gcttccttca catgctcttg ttttccagct	3211
gctttcctgg gggatcagac tgtgaagcag gaagacagat ataataaata tactgcatct	3271
ttttaagatg tgcaatttta ttctgaggaa acataaatta tgttttgtat tatatgactt	3331
taagagccca cattagggttt tatgattcat ttgccagggtt tttaaatggt ttcacaaaac	3391
tgttacggga cttcaactag aaataaaatg gtgtaaataa agacottgot atctctaaat	3451
tatggatggt aaagatttga aatgttttgt actttgatta tttttatttc ttatactctg	3511
ttttctttta tattgatatc ttgccacat tttaaataaa tgtacttttg aacttagaaa	3571
aaaaaaaa	3580





<210> 12  
 <211> 1218  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> (222)..(839)

<400> 12

```

atttggccct cgaggccaag aattcggcac gagcaacaat gactaatcaa actaacctca      60
aaacaaatga taaccataca caacactaaa ggacgaacct gatctcttat actagtatcc      120
ttaatcattt ttattgccac aactaacctc ctcggactcc tgccttactc atttacacca      180
accacccaac tatctataaa cctagccatg gccatcccct t      atg agc ggc cgc      233
                                   Met Ser Gly Arg
                                   1

agt ggt tat agg ctt tcg ctc ttg ctg ttg gcg gcg ctg ggc agc ggc      281
Ser Gly Tyr Arg Leu Ser Leu Leu Leu Leu Ala Ala Leu Gly Ser Gly
   5                10                15                20

cgc gcg gag cgc gac tgc cga gtg agc agc ttc cga gtc aag gag aac      329
Arg Ala Glu Arg Asp Cys Arg Val Ser Ser Phe Arg Val Lys Glu Asn
                25                30                35

ttc gac aag gct cgc ttc tct ggg acc tgg tac gcc atg gcc aag aag      377
Phe Asp Lys Ala Arg Phe Ser Gly Thr Trp Tyr Ala Met Ala Lys Lys
                40                45                50

gac ccc gag ggc ctc ttt ctg cag gac aac atc gtc gcg gag ttc tcc      425
Asp Pro Glu Gly Leu Phe Leu Gln Asp Asn Ile Val Ala Glu Phe Ser
                55                60                65

gtg gac gag acc ggc cag atg agc gcc aca gcc aag ggc cga gtc cgt      473
Val Asp Glu Thr Gly Gln Met Ser Ala Thr Ala Lys Gly Arg Val Arg
                70                75                80

ctt ttg aat aac tgg gac gtg tgc gca gac atg gtg ggc acc ttc aca      521
Leu Leu Asn Asn Trp Asp Val Cys Ala Asp Met Val Gly Thr Phe Thr
                85                90                95                100

gac acc gag gac cct gcc aag ttc aag atg aag tac tgg ggc gta gcc      569
Asp Thr Glu Asp Pro Ala Lys Phe Lys Met Lys Tyr Trp Gly Val Ala
                105                110                115

tcc ttt ctc cag aaa gga aat gat gac cac tgg atc gtc gac aca gac      617
Ser Phe Leu Gln Lys Gly Asn Asp Asp His Trp Ile Val Asp Thr Asp
                120                125                130

tac gac acg tat gcc gtg cag tac tcc tgc cgc ctc ctg aac ctc gat      665
Tyr Asp Thr Tyr Ala Val Gln Tyr Ser Cys Arg Leu Leu Asn Leu Asp
                135                140                145

```

```

ggc acc tgt gct gac agc tac tcc ttc gtg ttt tcc cgg gac ccc aac      713
Gly Thr Cys Ala Asp Ser Tyr Ser Phe Val Phe Ser Arg Asp Pro Asn
    150                      155                      160

ggc ctg ccc cca gaa gcg cag aag att gta agg cag cgg cag gag gag      761
Gly Leu Pro Pro Glu Ala Gln Lys Ile Val Arg Gln Arg Gln Glu Glu
    165                      170                      175                      180

ctg tgc ctg gcc agg cag tac agg ctg atc gtc cac aac ggt tac tgc      809
Leu Cys Leu Ala Arg Gln Tyr Arg Leu Ile Val His Asn Gly Tyr Cys
                      185                      190                      195

gat ggc aga tca gaa aga aac ctt ttg tag c aatatcaaga atctagtttc      860
Asp Gly Arg Ser Glu Arg Asn Leu Leu *
                      200                      205

atctgagaac ttctgattag ctctcagtct tcagctctat ttatcttagg agtttaattt      920

gcccttctct ccccatcttc cctcagttcc cataaaacct tcattacaca taaagataca      980

cgtgggggtc agtgaatctg cttgcctttc ctgaaagttt ctgggggtta agattccaga    1040

ctctgattca ttaaactata gtcaccgtgt ctggaaaaaa taaaaaaaaa tctctgccga    1100

ttcttggtctg aaggccaatt cctattgtgg tgcataaatc gcatatgtca tagtgtttct    1160

gggtgatgta ttccgtaaat tcaacacaac agccgaacct agtgaaggct ggtgccta      1218

<210> 13
<211> 4332
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (26)..(2539)

<400> 13
cagctcgggc gccccaccg gcccc atg gca gcc ccc ggc gcc cca gct gag      52
Met Ala Ala Pro Gly Ala Pro Ala Glu
                      1                      5

tac ggc tac atc cgg acc gtc ctg ggc cag cag atc ctg gga caa ctg      100
Tyr Gly Tyr Ile Arg Thr Val Leu Gly Gln Gln Ile Leu Gly Gln Leu
    10                      15                      20                      25

gac agc tcc agc ctg gcg ctg ccc tcc gag gcc aag ctg aag ctg gcg      148
Asp Ser Ser Ser Leu Ala Leu Pro Ser Glu Ala Lys Leu Lys Leu Ala
                      30                      35                      40

ggg agc agc ggc cgc ggc ggc cag aca gtc aag agc ctg cgg atc cag      196
Gly Ser Ser Gly Arg Gly Gly Gln Thr Val Lys Ser Leu Arg Ile Gln
                      45                      50                      55

```







730	735	740	745	
gaa act aca gcc tct gcc tgt tac aca ttg aac aac ata atc caa aac				2308
Glu Thr Thr Ala Ser Ala Cys Tyr Thr Leu Asn Asn Ile Ile Gln Asn	750	755	760	
agt tac cag aat gca cgc gac ctt cta aac acc ggg ggc atc cag aaa				2356
Ser Tyr Gln Asn Ala Arg Asp Leu Leu Asn Thr Gly Gly Ile Gln Lys	765	770	775	
att atg gcc att agt gca ggc gat gcc tat gcc tcc aac aaa gca agt				2404
Ile Met Ala Ile Ser Ala Gly Asp Ala Tyr Ala Ser Asn Lys Ala Ser	780	785	790	
aaa gct gct tcc gtc ctt ctg tat tct ctg tgg gca cac acg gaa ctg				2452
Lys Ala Ala Ser Val Leu Leu Tyr Ser Leu Trp Ala His Thr Glu Leu	795	800	805	
cat cat gcc tac aag aag gct cag ttt aag aag aca gat ttt gtc aac				2500
His His Ala Tyr Lys Lys Ala Gln Phe Lys Lys Thr Asp Phe Val Asn	810	815	820	825
agc cgg act gcc aaa gcc tac cac tcc ctt aaa gac tga ggaaaatgac				2549
Ser Arg Thr Ala Lys Ala Tyr His Ser Leu Lys Asp *	830	835		
aaagtattct cggctgcaaa aatccccaaa ggaaaacacc tatttttcta ctaccagcc				2609
caagaaacct caaaagcatg ccttgtttct atccttctct atttcogtgg tccctgaat				2669
ccagaaaaca aatagaacat aattttatga gtcttccaga agaccttgc aagtttgcca				2729
ccagtagata ccggccacag gctcgacaaa tagtgggtctt tggtattagg gcttatggta				2789
gatggccttc tggaatcaaa atgtgaattc atgtggaagg gacattaatc caataaataa				2849
ggaaagaagc tgttgcatta ctgggatttt aaaagtttga tttacattta tattcctttt				2909
ctgggtccca tgttttgtca ctcatgtgca cattgcttcg ccattgggcc tccagtgtat				2969
tgttctgcag tgttgaaaca gaatggaaat gacaagaaat atctgcagtt atccaggaga				3029
aagtataatg gcaaaattat tggtttcttt ctttactttg tgcttgtttt tatccccctg				3089
gttggtttttc tctgattttt aaataaactt aagaaattta gattacagag tatgcatgac				3149
tgtaagaaaa agaaattgag aggaagtgat catagcaaat taaagaagtc ttttcctccc				3209
agaacttaaa gtaaaataaa aaataaataa ataaataaaa tcttttccac agagaaaggc				3269
aactgtgatg ataaaattta acgttcccc aaacactgag tcaatgagat ttttctcagg				3329
agatacttta cctataacaa cgccgttaaa tccaaatctc ttctaaacga tggcattcta				3389
tgtaatgcct ttcttggaact tttttggcca ctgccctgga ctagtgaaag aatggactct				3449
atctttatct gcaagaggaa ctaaggcctt ctatcagact gcctggccag cctggggcac				3509

tgaaaatacg gctcatgtta atgagttaca ttatcagcca gccagcctt gccaccatt 3569  
taagaaatat cacagagcca ctagatctca tatgatcttc ttcaagccat tattttaact 3629  
caagaaaact ctagagaaga aaagtgaaga agtcatgttg aagaagatgt aagaatgtgt 3689  
caagaccatc cagaaatgat atgagaaata ctgatatttt aaatggttga catcatccag 3749  
cgaaatgaat ctacattaaa tgttgtttta actgcgctat gattaaaacc attcatatag 3809  
agttagtctt tacaactact attctgttat tttttttttt aatctgacaa catttgcct 3869  
aagtaagata agcaaaaaa ttcttcaact ccttttggca agaaaactgt aacagaaaat 3929  
aaattttgaa tgtgtactta agtctttatt atatttgaag caattttttt tcaattttta 3989  
aagctgaatg aagacaactt aggttgctaa cctagttcaa aatgaaatta tttagatacc 4049  
aattttttaa atactggaga gaatttatat gtctttttcc agagttctga ttataagcat 4109  
ttggagtgca tttattcctc cagataataa atgtgtgttc agaacttttt gtgtttttta 4169  
aggcattaat aaagccttcg ataataataa atacaaaaa aaagggcggc cgctctagag 4229  
tatccctcga ggggcccaag cttacgcgta ccagccttc ttgtacaaag tggtcctat 4289  
agtgagtcgt atataagcag cgacgttcaa agagactaca aat 4332

<210> 14  
<211> 1045  
<212> DNA  
<213> Homo sapiens  
  
<220>  
<221> CDS  
<222> (51) .. (722)

<400> 14  
ctgcaccata ttccagaaca actgagcaaa caagagaggt acaccgacaa atg gca 56  
Met Ala  
1  
  
gcg acg cac aca ccc cct ctg gcg tca ccg ggc atg gcg cgc atc tgc 104  
Ala Thr His Thr Pro Pro Leu Ala Ser Pro Gly Met Ala Arg Ile Cys  
5 10 15  
  
ctt tac ggg gat tta caa cga ttt ggt cgc cgc atc gac ctt cgt gtg 152  
Leu Tyr Gly Asp Leu Gln Arg Phe Gly Arg Arg Ile Asp Leu Arg Val  
20 25 30  
  
aaa acg ggg gct gaa gcc atc cgg gca ctg gcc aca cag ctc ccg gcg 200  
Lys Thr Gly Ala Glu Ala Ile Arg Ala Leu Ala Thr Gln Leu Pro Ala  
35 40 45 50



ttt cgt cag aaa ctg agc gac ggc tgg tat cag gta cgg att gcc ggg	248
Phe Arg Gln Lys Leu Ser Asp Gly Trp Tyr Gln Val Arg Ile Ala Gly	
55 60 65	
cgg gac gtc agc acg tcc ggg tta acg gcg cag tta cat gag act ctg	296
Arg Asp Val Ser Thr Ser Gly Leu Thr Ala Gln Leu His Glu Thr Leu	
70 75 80	
cct gat ggc gct gta att cat att gtt ccc aga gtc gcc ggg gcc aag	344
Pro Asp Gly Ala Val Ile His Ile Val Pro Arg Val Ala Gly Ala Lys	
85 90 95	
tca ggt ggc gta ttc cag att gtc ctg ggg gct gcc gcc att gcc gga	392
Ser Gly Gly Val Phe Gln Ile Val Leu Gly Ala Ala Ala Ile Ala Gly	
100 105 110	
tca ttc ttt acc gcc gga gcc acc ctt gca gca tgg ggg gca gcc att	440
Ser Phe Phe Thr Ala Gly Ala Thr Leu Ala Ala Trp Gly Ala Ala Ile	
115 120 125 130	
ggg gcc ggt ggt atg acc ggc atc ctg ttt tct ctc ggt gcc agt atg	488
Gly Ala Gly Gly Met Thr Gly Ile Leu Phe Ser Leu Gly Ala Ser Met	
135 140 145	
gtg ctc ggt ggt gtg gcg cag atg ctg gca ccg aaa gcc aga act ccc	536
Val Leu Gly Gly Val Ala Gln Met Leu Ala Pro Lys Ala Arg Thr Pro	
150 155 160	
cgt ata cag aca acg gat aac ggt aag cag aac acc tat ttc tcc tca	584
Arg Ile Gln Thr Thr Asp Asn Gly Lys Gln Asn Thr Tyr Phe Ser Ser	
165 170 175	
ctg gat aac atg gtt gcc cag ggc aat gtt ctg cct gtt ctg tac ggg	632
Leu Asp Asn Met Val Ala Gln Gly Asn Val Leu Pro Val Leu Tyr Gly	
180 185 190	
gaa atg cgc gtg ggg tca cgc gtg gtt tct cag gag atc agc acg gca	680
Glu Met Arg Val Gly Ser Arg Val Val Ser Gln Glu Ile Ser Thr Ala	
195 200 205 210	
gac gaa ggg gac ggt ggt cag gtt gtg gtg att ggt cgc tga tgcaaaa	729
Asp Glu Gly Asp Gly Gly Gln Val Val Val Ile Gly Arg *	
215 220	
tgttttatgt gaaaccgcct gcgggcgggt ttgtcattta tggagcgtga ggaatgggta	789
aaggaagcag taaggggcat accccgcgcg aagcgaagga caacctgaag tccacgcagt	849
tgctgagtgt gatcgatgcc atcagcgaag ggccgattga aggtccgggtg gatggcttaa	909
aaagcgtgct gctgaacagt acgccggtgc tggacactga ggggaatacc aacatatccg	969
gtgtcacggg ggcgttcccg gctgttgagc aggatcatac tcccccccaa aggatttggg	1029
tcctcccccc cccac	1045

<210> 15  
 <211> 797  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> (101)..(754)

<400> 15

tgga	cggcga	tcagtcgact	agcgtttaa	aa	cttaagcttg	gtaccgagct	cggatccact	60
agtccagtg	ggtggaattc	aacaggagga	gaaattaggt	atg	gat	gac	ttg act	115
				Met	Asp	Asp	Leu Thr	
				1			5	
aag	ttg	gta	cag	gaa	cag	aaa	cct aaa ggc agt	163
Lys	Leu	Val	Gln	Glu	Gln	Lys	Pro Lys Gly Ser Gln Arg Ser Arg Lys	
			10				15 20	
aga	ggc	cat	acg	gct	tca	gaa	tct gat gaa cag cag tgg cct gag gaa	211
Arg	Gly	His	Thr	Ala	Ser	Glu	Ser Asp Glu Gln Gln Trp Pro Glu Glu	
			25				30 35	
aag	agg	ctc	aaa	gaa	gat	ata	tta gaa aat gaa gat gaa cag aat agt	259
Lys	Arg	Leu	Lys	Glu	Asp	Ile	Leu Glu Asn Glu Asp Glu Gln Asn Ser	
			40				45 50	
ccg	cca	aaa	aag	ggt	aaa	aga	ggc cga cca cca aaa cct ctt ggt gga	307
Pro	Pro	Lys	Lys	Gly	Lys	Arg	Gly Arg Pro Pro Lys Pro Leu Gly Gly	
			55				60 65	
ggt	aca	cca	aaa	gaa	gag	cca	aca atg aaa act tct aaa aaa gga agc	355
Gly	Thr	Pro	Lys	Glu	Glu	Pro	Thr Met Lys Thr Ser Lys Lys Gly Ser	
			70				75 80 85	
aaa	aaa	aaa	tct	gga	cct	cca	gca cca gag gag gag gaa gaa gaa gaa	403
Lys	Lys	Lys	Ser	Gly	Pro	Pro	Ala Pro Glu Glu Glu Glu Glu Glu	
			90				95 100	
aga	caa	agt	gga	aat	acg	gaa	cag aag tcc aaa agc aaa cag cac cga	451
Arg	Gln	Ser	Gly	Asn	Thr	Glu	Gln Lys Ser Lys Ser Lys Gln His Arg	
			105				110 115	
gtg	tca	agg	aga	gca	cag	cag	aga gca gaa tct cct gaa tct agt gca	499
Val	Ser	Arg	Arg	Ala	Gln	Gln	Arg Ala Glu Ser Pro Glu Ser Ser Ala	
			120				125 130	
att	gaa	tcc	aca	cag	tcc	aca	cca cag aaa gga cga gga aga cca tca	547
Ile	Glu	Ser	Thr	Gln	Ser	Thr	Pro Gln Lys Gly Arg Gly Arg Pro Ser	
			135				140 145	
aaa	acg	cca	tca	cca	tca	caa	cca aaa aaa aat gtc cgt gta gga cgc	595
Lys	Thr	Pro	Ser	Pro	Ser	Gln	Pro Lys Lys Asn Val Arg Val Gly Arg	

150	155	160	165	
tcc aaa caa gca gct act aag gaa aat gat tca agt gaa gaa gta gat				643
Ser Lys Gln Ala Ala Thr Lys Glu Asn Asp Ser Ser Glu Glu Val Asp				
	170	175	180	
gtg ttt cag ggt agc tct cct gtc gat gat att cca cag gaa gaa aca				691
Val Phe Gln Gly Ser Ser Pro Val Asp Asp Ile Pro Gln Glu Glu Thr				
	185	190	195	
gag gag gag gaa gtt tct aca gta aat gta cgg cgg cga agt gct aaa				739
Glu Glu Glu Glu Val Ser Thr Val Asn Val Arg Arg Arg Ser Ala Lys				
	200	205	210	
agg gaa cgg cga tga acaaatgtaa ttaataactt tctctgtgaa agctttggaa				794
Arg Glu Arg Arg *				
	215			
aaa				797
<210> 16				
<211> 1598				
<212> DNA				
<213> Homo sapiens				
<220>				
<221> CDS				
<222> (244) .. (1362)				
<400> 16				
atttggccct cgaggccaag aattcggcac gatgaagatc gagcgaggag cccaagagcg				60
agcgcgacgc acgaagctcg agccgcctcc gccgcgcgac cccacctcgg ccgccgccgc				120
ctgcgccgcg agatccgccc cggcctcccc gagagcgagc cccggccgcc gcgaccacca				180
gccgcgctaa ccgccgacca accgccaccg aggcgcctga gcgagagcag aggaggagga				240
ggc atg agt gag gcg ggc gag gcc acc acc acc acc acc acc acc ctc				288
Met Ser Glu Ala Gly Glu Ala Thr Thr Thr Thr Thr Thr Thr Leu				
1 5 10 15				
ccg cag gct ccg acg gag gcg gcc gcc gcg gct ccc cag gac ccc gcg				336
Pro Gln Ala Pro Thr Glu Ala Ala Ala Ala Ala Pro Gln Asp Pro Ala				
	20	25	30	
ccc aag agc ccg gtg ggc agc ggt gcg ccc cag gcc gcg gcc ccg gcg				384
Pro Lys Ser Pro Val Gly Ser Gly Ala Pro Gln Ala Ala Ala Pro Ala				
	35	40	45	
ccc gcc gcc cac gtc gca gga aac ccc ggt ggg gac gcg gcc ccc gca				432
Pro Ala Ala His Val Ala Gly Asn Pro Gly Gly Asp Ala Ala Pro Ala				
	50	55	60	

gcc acg ggc acc gcg gcc gcc gcc tct tta gcc acc gcc gcc ggc agc Ala Thr Gly Thr Ala Ala Ala Ala Ser Leu Ala Thr Ala Ala Gly Ser 65 70 75	480
gaa gac gcg gag aaa aaa gtt ctc gcc acc aaa gtc ctt ggc act gtc Glu Asp Ala Glu Lys Lys Val Leu Ala Thr Lys Val Leu Gly Thr Val 80 85 90 95	528
aaa tgg ttc aac gtc aga aat gga tat gga ttt ata aat cga aat gac Lys Trp Phe Asn Val Arg Asn Gly Tyr Gly Phe Ile Asn Arg Asn Asp 100 105 110	576
acc aaa gaa gat gta ttt gta cat cag act gcc atc aag aag aat aac Thr Lys Glu Asp Val Phe Val His Gln Thr Ala Ile Lys Lys Asn Asn 115 120 125	624
cca cgg aaa tat ctg cgc agt gta gga gat gga gaa act gta gag ttt Pro Arg Lys Tyr Leu Arg Ser Val Gly Asp Gly Glu Thr Val Glu Phe 130 135 140	672
gat gtg gtt gaa gga gag aag ggt gca gaa gct gcc aat gtg act ggc Asp Val Val Glu Gly Glu Lys Gly Ala Glu Ala Ala Asn Val Thr Gly 145 150 155	720
ccg gat gga gtt cct gtg gaa ggg agt cgt tac gct gca gat cgg cgc Pro Asp Gly Val Pro Val Glu Gly Ser Arg Tyr Ala Ala Asp Arg Arg 160 165 170 175	768
cgt tac aga cgt ggc tac tat gga agg cgc cgt ggc cct ccc cgg aat Arg Tyr Arg Arg Gly Tyr Tyr Gly Arg Arg Arg Gly Pro Pro Arg Asn 180 185 190	816
tac gct ggg gag gag gag gag gaa ggg agc ggc agc agt gaa gga ttt Tyr Ala Gly Glu Glu Glu Glu Glu Gly Ser Gly Ser Ser Glu Gly Phe 195 200 205	864
gac ccc cct gcc act gat agg cag ttc tct ggg gcc cgg aat cag ctg Asp Pro Pro Ala Thr Asp Arg Gln Phe Ser Gly Ala Arg Asn Gln Leu 210 215 220	912
cgc cgc ccc cag tat cgc cct cag tac cgg cag cgg cgg ttc ccg cct Arg Arg Pro Gln Tyr Arg Pro Gln Tyr Arg Gln Arg Arg Phe Pro Pro 225 230 235	960
tac cac gtg gga cag acc ttt gac cgt cgc tca cgg gtc tta ccc cat Tyr His Val Gly Gln Thr Phe Asp Arg Arg Ser Arg Val Leu Pro His 240 245 250 255	1008
ccc aac aga ata cag gct ggt gag att gga gag atg aag gat gga gtc Pro Asn Arg Ile Gln Ala Gly Glu Ile Gly Glu Met Lys Asp Gly Val 260 265 270	1056
cca gag gga gca caa ctt cag gga ccg gtt cat cga aat cca act tac Pro Glu Gly Ala Gln Leu Gln Gly Pro Val His Arg Asn Pro Thr Tyr 275 280 285	1104
cgc cca agg tac cgt agc agg gga cct cct cgc cca cga cct gcc cca	1152





240	245	250	255	
gtt cgc cgt gga tac cgg cgt ccc tac aat tac cgg cgt cgc ccg cgt				1056
Val Arg Arg Gly Tyr Arg Arg Pro Tyr Asn Tyr Arg Arg Arg Pro Arg				
	260	265	270	
cct cct aac gct cct tca caa gat ggc aaa gag gcc aag gca ggt gaa				1104
Pro Pro Asn Ala Pro Ser Gln Asp Gly Lys Glu Ala Lys Ala Gly Glu				
	275	280	285	
gca cca act gag aac cct gct cca ccc acc cag cag agc agt gct gag				1152
Ala Pro Thr Glu Asn Pro Ala Pro Pro Thr Gln Gln Ser Ser Ala Glu				
	290	295	300	
taa cacc aggctcctca ggcaccttca ccatacggcag gtgacctaaa gaattaatga				1209
*				
ccattcagaa ataaagcaaa aagcaggcca caaccttaac caacacccaaa gaaacatcca				1269
agcaataaag tggaagacta accaagattt ggacattgga atgtttactg ttattcttta				1329
agaaacaact acaaaaagaa aatgtcaaca aatttttcca gcaagctgag aacctggaat				1389
tc				1391
<210>	18			
<211>	787			
<212>	DNA			
<213>	Homo sapiens			
<220>				
<221>	CDS			
<222>	(84) .. (395)			
<400>	18			
gggagagcgg cggggccgag agcgtgactc gcccgctccg cgctgcttcc cccgcgcgcg				60
ctccccgcgc cgctcgcgca gcc		atg tcc gag gag aag ccc aag gag ggt		110
		Met Ser Glu Glu Lys Pro Lys Glu Gly		
		1 5		
gtg aag aca gag aat gac cac atc aac ctg aag gtg gcc ggg cag gac				158
Val Lys Thr Glu Asn Asp His Ile Asn Leu Lys Val Ala Gly Gln Asp				
10 15 20 25				
ggc tcc gtg gtg cag ttc aag atc aag agg cac acg ccg ctg agc aag				206
Gly Ser Val Val Gln Phe Lys Ile Lys Arg His Thr Pro Leu Ser Lys				
30 35 40				
ctg atg aag gcc tac tgc gag agg cag ggc ttg tca atg agg cag atc				254
Leu Met Lys Ala Tyr Cys Glu Arg Gln Gly Leu Ser Met Arg Gln Ile				
45 50 55				

```

aga ttc agg ttc gac ggg cag cca atc aat gaa act gac act cca gca      302
Arg Phe Arg Phe Asp Gly Gln Pro Ile Asn Glu Thr Asp Thr Pro Ala
      60              65              70

cag ctg gag atg gag gac gag gac acc atc gac gtg ttc cag cag cag      350
Gln Leu Glu Met Glu Asp Glu Asp Thr Ile Asp Val Phe Gln Gln Gln
      75              80              85

acg gga ggt gtg ccg gag agc agc ctg gca ggg cac agt ttc tag agg      398
Thr Gly Gly Val Pro Glu Ser Ser Leu Ala Gly His Ser Phe *
      90              95              100

gcccgcccc agcccgggcc gtccatcctc gcattgctgt tgaatggtga gcacgtgacc      458

atgccgacca caaaggtgtc tgcggaaact cgaggacatt caccacgatg attttcctct      518

ctttgatgta cttcaagtgc aactcaaaac tatactctgca gggatgaatc tgtaacttaa      578

attggggccaa tcagaattgt tatctttgtt caggtaaaat gagttgcaag atattgtggg      638

tacttttgtg tgctcatttg tgttttcccc cctcctaca acattttttt aacccccaaaa      698

ttatagcctg aatgttcgct ttagtctgg ccagggatct gactcctgag ttggttgctt      758

ctccccctgct cactccagtc acatagaga                                          787

```

```

<210> 19
<211> 1576
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> CDS
<222> (43)..(1293)

```

```

<400> 19
atttggccct cgaggccaag aattcggcac gaggccaggg cc      atg gcg gag gag      54
                                     Met Ala Glu Glu
                                     1

cag ggc cgg gaa cgg gac tcg gtt ccc aag ccg tcg gtg ctg ttc ctc      102
Gln Gly Arg Glu Arg Asp Ser Val Pro Lys Pro Ser Val Leu Phe Leu
      5              10              15              20

cac cca gac ctg ggc gtg ggc ggc gct gag cgg ctg gtg ttg gac gcg      150
His Pro Asp Leu Gly Val Gly Gly Ala Glu Arg Leu Val Leu Asp Ala
              25              30              35

gcg ctg gcg ctg cag gcg cgc ggg tgt agc gtg aag atc tgg aca gcg      198
Ala Leu Ala Leu Gln Ala Arg Gly Cys Ser Val Lys Ile Trp Thr Ala
              40              45              50

cac tac gac ccg ggc cac tgt ttc gcc gag agc cgc gag cta ccg gtg      246
His Tyr Asp Pro Gly His Cys Phe Ala Glu Ser Arg Glu Leu Pro Val

```



55	60	65	
cgc tgt gcc ggg gac tgg ctg ccg cga ggc ctg ggc tgg ggc ggc cgc Arg Cys Ala Gly Asp Trp Leu Pro Arg Gly Leu Gly Trp Gly Gly Arg 70 75 80			294
ggc gcc gcc gtc tgc gcc tac gtg cgc atg gtt ttc ctg gcg ctc tac Gly Ala Ala Val Cys Ala Tyr Val Arg Met Val Phe Leu Ala Leu Tyr 85 90 95 100			342
gtg ctg ttc ctc gcc gac gag gag ttc gac gtg gta gtg tgc gac cag Val Leu Phe Leu Ala Asp Glu Glu Phe Asp Val Val Val Cys Asp Gln 105 110 115			390
gtg tct gcc tgt atc cca gtg ttc agg ctg gct aga cgg cgg aag aag Val Ser Ala Cys Ile Pro Val Phe Arg Leu Ala Arg Arg Arg Lys Lys 120 125 130			438
atc cta ttt tac tgt cac ttc cca gat ctg ctt ctc acc aag aga gat Ile Leu Phe Tyr Cys His Phe Pro Asp Leu Leu Leu Thr Lys Arg Asp 135 140 145			486
tct ttt ctt aaa cga cta tac agg gcc cca att gac tgg ata gag gaa Ser Phe Leu Lys Arg Leu Tyr Arg Ala Pro Ile Asp Trp Ile Glu Glu 150 155 160			534
tac acc aca ggc atg gca gac tgc atc tta gtc aac agc cag ttc aca Tyr Thr Thr Gly Met Ala Asp Cys Ile Leu Val Asn Ser Gln Phe Thr 165 170 175 180			582
gct gct gtt ttt aag gaa aca ttc aag tcc ctg tct cac ata gac cct Ala Ala Val Phe Lys Glu Thr Phe Lys Ser Leu Ser His Ile Asp Pro 185 190 195			630
gat gtc ctc tat cca tct cta aat gtc acc agc ttt gac tca gtt gtt Asp Val Leu Tyr Pro Ser Leu Asn Val Thr Ser Phe Asp Ser Val Val 200 205 210			678
cct gaa aag ctg gat gac cta gtc ccc aag ggg aaa aaa ttc ctg ctg Pro Glu Lys Leu Asp Asp Leu Val Pro Lys Gly Lys Lys Phe Leu Leu 215 220 225			726
ctc tcc atc aac aga tac gaa agg aag aaa aat ctg act ttg gca ctg Leu Ser Ile Asn Arg Tyr Glu Arg Lys Lys Asn Leu Thr Leu Ala Leu 230 235 240			774
gaa gcc cta gta cag ctg cgt gga aga ttg aca tcc caa gat tgg gag Glu Ala Leu Val Gln Leu Arg Gly Arg Leu Thr Ser Gln Asp Trp Glu 245 250 255 260			822
agg gtt cat ctg atc gtg gca ggt ggt tat gac gag aga gtc ctg gag Arg Val His Leu Ile Val Ala Gly Gly Tyr Asp Glu Arg Val Leu Glu 265 270 275			870
aat gtg gaa cat tat cag gaa ttg aag aaa atg gtc caa cag tcc gac Asn Val Glu His Tyr Gln Glu Leu Lys Lys Met Val Gln Gln Ser Asp 280 285 290			918





155	160	165	
cac ctc atc ctg cgt cag agg act	cac acg ggg gag aag ccc tgc gac	1061	
His Leu Ile Leu Arg Gln Arg Thr	His Thr Gly Glu Lys Pro Cys Asp		
170	175	180	
tgc gcg gag tgc ggg aaa gct gct	ccc agg act cct gcc tcc tga	1106	
Cys Ala Glu Cys Gly Lys Ala Ala	Pro Arg Thr Pro Ala Ser *		
185	190	195	

<210> 21  
 <211> 4164  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> (2956) .. (3981)

<400> 21	
atgacatcgc tgcgcgaata tgccggttat cacggcggtg gcagcggatt tggagggcag	60
ttgcggtcgt ggaaccacc gagtgaaagt gtggatgcag ccctgttgcc caactttacc	120
cgtggcaatg cccgcgcaga cgatctggta cgcaataacg gctatgccgc caacgccatc	180
cagctgcac accgatcatat cgtcgggtct tttttccggc tcagtcacgc cccaagctgg	240
cgctatctgg gcatcgggga ggaagaagcc cgtgcctttt cccgcgaggt tgaagcggca	300
tggaaagagt ttgccgagga tgactgctgc tgcattgacg ttgagcgaaa acgcacgttt	360
accatgatga ttcgggaagg tgtggccatg cacgccttta acggtgaact gttcgttcag	420
gccacctggg ataccagttc gtcgcggctt ttccggacac agttccggat ggtcagccc	480
aagcgcacga gcaacccgaa caataccggc gacagccgga actgccgtgc cgggtgtgcag	540
attaatgaca gcggtgcggc gctgggatat tacgtcagcg aggacgggta tcctggctat	600
atgccgcaga aatggacatg gataccccgt gagttaccgc gcgggcgcgc ctcgttcatt	660
cacgtttttg aacccgtgga ggacgggcag actcgcgggtg caaatgtgtt ttacagcgtg	720
atggagcaga taaagatgct cgacacgctg cagaacacgc agctgcagag cgccattgtg	780
aaggcgatgt atgccgccac cattgagagt gagctggata cgcagtcagc gatggatttt	840
attctgggcg cgaacagtca ggagcagcgg gaaaggctga cggctggat tggtgaaatt	900
gccgcgtatt acgccgcagc gccgggtccg ctgggaggcg caaaagtacc gcacctgatg	960
ccgggtgact cactgaacct gcagacggct caggatacgg ataacggcta ctccgtgttt	1020



gcaggtaacc cggcatctga tgccgttaac gatttgctga acacaccagt ccgtaagctg	2760
gttgctggtg atggcaccac cgacggtgct gccgttggca ttcttgcggt tgctgctgac	2820
cagaccagca ccacgctgac gttctacaag tccggcacgt tccgttatga ggatgtgctc	2880
tggccggagg ctgccagcga cgagacgaaa aaacggaccg cgccgcctgt gcggtttttt	2940
ttacgggatt ttttt atg tcg atg tac aca acc gcc caa ctg ctg gcg gca	2991
Met Ser Met Tyr Thr Thr Ala Gln Leu Leu Ala Ala	
1 5 10	
aat gag cag aaa ttt aag ttt gat ccg ctg ttt ctg cgt ctc ttt ttc	3039
Asn Glu Gln Lys Phe Lys Phe Asp Pro Leu Phe Leu Arg Leu Phe Phe	
15 20 25	
cgt gag agc tat ccc ttc acc acg gag aaa gtc tat ctc tca caa att	3087
Arg Glu Ser Tyr Pro Phe Thr Thr Glu Lys Val Tyr Leu Ser Gln Ile	
30 35 40	
ccg gga ctg gta aac atg gcg ctg tac gtt tgc ccg att gtt tcc ggt	3135
Pro Gly Leu Val Asn Met Ala Leu Tyr Val Ser Pro Ile Val Ser Gly	
45 50 55 60	
gag gtt atc cgt tcc cgt ggc ggc tcc acc tct gaa ttt acg ccg gga	3183
Glu Val Ile Arg Ser Arg Gly Gly Ser Thr Ser Glu Phe Thr Pro Gly	
65 70 75	
tat gtc aag ccg aag cat gaa gtg aat ccg cag atg acc ctg cgt cgc	3231
Tyr Val Lys Pro Lys His Glu Val Asn Pro Gln Met Thr Leu Arg Arg	
80 85 90	
ctg ccg gat gaa gat ccg cag aat ctg gcg gac ccg gct tac cgc cgc	3279
Leu Pro Asp Glu Asp Pro Gln Asn Leu Ala Asp Pro Ala Tyr Arg Arg	
95 100 105	
cgt cgc atc atc atg cag aac atg cgt gac gaa gag ctg gcc att gct	3327
Arg Arg Ile Ile Met Gln Asn Met Arg Asp Glu Glu Leu Ala Ile Ala	
110 115 120	
cag gtc gaa gag atg cag gca gtt tct gcc gtg ctt aag ggc aaa tac	3375
Gln Val Glu Glu Met Gln Ala Val Ser Ala Val Leu Lys Gly Lys Tyr	
125 130 135 140	
acc atg acc ggt gaa gcc ttc gat ccg gtt gag gtg gat atg ggc cgc	3423
Thr Met Thr Gly Glu Ala Phe Asp Pro Val Glu Val Asp Met Gly Arg	
145 150 155	
agt gag gag aat aac atc acg cag tcc ggc ggc acg gag tgg agc aag	3471
Ser Glu Glu Asn Asn Ile Thr Gln Ser Gly Gly Thr Glu Trp Ser Lys	
160 165 170	
cgt gac aag tcc acg tat gac ccg acc gac gat atc gaa gcc tac gcg	3519
Arg Asp Lys Ser Thr Tyr Asp Pro Thr Asp Asp Ile Glu Ala Tyr Ala	
175 180 185	
ctg aac gcc agc ggt gtg gtg aat atc atc gtg ttc gat ccg aaa ggc	3567

```

Leu Asn Ala Ser Gly Val Val Asn Ile Ile Val Phe Asp Pro Lys Gly
190                               195                               200

tgg gcg ctg ttc cgt tcc ttc aaa gcc gtc aag gag aag ctg gat acc 3615
Trp Ala Leu Phe Arg Ser Phe Lys Ala Val Lys Glu Lys Leu Asp Thr
205                               210                               215                               220

cgt cgt ggc tct aat tcc gag ctg gag aca gcg gtg aaa gac ctg ggc 3663
Arg Arg Gly Ser Asn Ser Glu Leu Glu Thr Ala Val Lys Asp Leu Gly
225                               230                               235

aaa gcg gtg tcc tat aag ggg atg tat ggc gat gtg gcc atc gtc gtg 3711
Lys Ala Val Ser Tyr Lys Gly Met Tyr Gly Asp Val Ala Ile Val Val
240                               245                               250

tat tcc gga cag tac gtg gaa aac ggc gtc aaa aag aac ttc ctg ccg 3759
Tyr Ser Gly Gln Tyr Val Glu Asn Gly Val Lys Lys Asn Phe Leu Pro
255                               260                               265

gac aac acg atg gtg ctg ggg aac act cag gca cgc ggt ctg cgc acc 3807
Asp Asn Thr Met Val Leu Gly Asn Thr Gln Ala Arg Gly Leu Arg Thr
270                               275                               280

tat ggc tgc att cag gat gcg gac gca cag cgc gaa ggc att aac gcc 3855
Tyr Gly Cys Ile Gln Asp Ala Asp Ala Gln Arg Glu Gly Ile Asn Ala
285                               290                               295                               300

tct gcc cgt tac ccg aaa aac tgg gtg acc acc ggc gat ccg gcg cgt 3903
Ser Ala Arg Tyr Pro Lys Asn Trp Val Thr Thr Gly Asp Pro Ala Arg
305                               310                               315

gag ttc acc atg att cag tca gca ccg ctg atg ctg ctg gct gac cct 3951
Glu Phe Thr Met Ile Gln Ser Ala Pro Leu Met Leu Leu Ala Asp Pro
320                               325                               330

gat gag ttc gtg tcc gta caa ctg gcg taa t catggccctt cggggccatt 4002
Asp Glu Phe Val Ser Val Gln Leu Ala *
335                               340

gtttctctgt ggaggagtcc atgacgaaag atgaactgat tgcccgctctt cgctcgctgg 4062

gtgaacaact tgaaccgga tgtcaagcct gacggggggac gaaagaagaa ctggcgctcc 4122

gttggtggca agagcttgaa agaggagctt gatgacacga gg 4164

```

<210> 22  
 <211> 541  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> (192)..(425)





tat ggc tgg tac atc gtc ttc agc tgc atc ctt ctc tac gtg gtc ttt Tyr Gly Trp Tyr Ile Val Phe Ser Cys Ile Leu Leu Tyr Val Val Phe 35 40 45	203
cag aag ctt tcc gcc cgg cta aga gcc ttg agg cag agg cag ctg gac Gln Lys Leu Ser Ala Arg Leu Arg Ala Leu Arg Gln Arg Gln Leu Asp 50 55 60	251
cga gct gcg gct gct gtg gaa cct gat gtt gtt gtt aaa cga caa gaa Arg Ala Ala Ala Ala Val Glu Pro Asp Val Val Val Lys Arg Gln Glu 65 70 75 80	299
gct tta gca gct gct cga ctg aaa atg caa gaa gaa cta aat gcg caa Ala Leu Ala Ala Ala Arg Leu Lys Met Gln Glu Glu Leu Asn Ala Gln 85 90 95	347
gtt gaa aag cat aag gaa aaa ctg aaa caa ctt gaa gaa gaa aaa agg Val Glu Lys His Lys Glu Lys Leu Lys Gln Leu Glu Glu Glu Lys Arg 100 105 110	395
aga cag aag att gaa atg tgg gac agc atg caa gaa gga aaa agt tac Arg Gln Lys Ile Glu Met Trp Asp Ser Met Gln Glu Gly Lys Ser Tyr 115 120 125	443
aaa gga aat gca aag aag ccc cag gag gaa gac agt cct ggg cct tcc Lys Gly Asn Ala Lys Lys Pro Gln Glu Glu Asp Ser Pro Gly Pro Ser 130 135 140	491
act tca tct gtc ctg aaa cgg aaa tcg gac aga aag cct ttg cgg gga Thr Ser Ser Val Leu Lys Arg Lys Ser Asp Arg Lys Pro Leu Arg Gly 145 150 155 160	539
gga ggt tat aac ccg ttg tct ggt gaa gga ggc gga gct tgc tcc tgg Gly Gly Tyr Asn Pro Leu Ser Gly Glu Gly Gly Gly Ala Cys Ser Trp 165 170 175	587
aga cct gga cgc aga ggc ccg tca tct ggc gga tga ggct aagaatcttg Arg Pro Gly Arg Arg Gly Pro Ser Ser Gly Gly *	637
180 185	
ttagtgtcac ttttgacatt agcaagatga acccttaacc ctcgattcaa ttgccttacg	697
cacgctttttc acagtgacta gccaaagggga ggtgggggttg atttctgttc ctaactacac	757
ctgcatatgt cagggctcca gtcagcaaaa ggtatagatg ttgcctctag gcatgaggtc	817
attggtcaca ttctacttgg agacagtgat tgcattcatt gatttcatgg ttaattgcta	877
gttggttaggt aaaggcctct agatgattag caatcttgat aaaagaggcc tagtaatgtt	937
cttttgaggt tagaaatcct tgctgctagg acagtctctg tgacagggtg cgttgaatga	997
tgtcttcctt atcaatggtg agcccaccag tgaggattac tgatgtggac agttgatggg	1057
gtttgtttct gtatatattat ttttatgtac agaactttgt aaaaacgaaa ctatttaaaa	1117
aacaagaata acatttttag catctttatt taatgtttga tagacaaatt gaagtccata	1177



Ala	Leu	Lys	Lys	Cys	Phe	Pro	Val	Val	Glu	Glu	Gln	Gln	Gly	Leu	Trp	
75					80					85					90	
cag	agt	gcc	ctg	cgg	gac	tgc	cag	ccc	ctc	ctg	tcc	tcc	ctc	agc	aac	519
Gln	Ser	Ala	Leu	Arg	Asp	Cys	Gln	Pro	Leu	Leu	Ser	Ser	Leu	Ser	Asn	
				95					100					105		
ctg	gcg	gaa	cag	ctg	cag	gcc	gca	cag	aac	ctg	cgg	ttt	gag	gat	gtg	567
Leu	Ala	Glu	Gln	Leu	Gln	Ala	Ala	Gln	Asn	Leu	Arg	Phe	Glu	Asp	Val	
			110					115					120			
cgg	gcg	ctt	cgg	gcc	ttc	cca	gat	tta	aaa	gag	cgg	ctg	agg	cgt	aag	615
Pro	Ala	Leu	Arg	Ala	Phe	Pro	Asp	Leu	Lys	Glu	Arg	Leu	Arg	Arg	Lys	
		125					130					135				
cag	ctg	gtg	gct	ggg	gac	atc	gtc	ctg	gac	aag	cta	ggg	gaa	agg	cta	663
Gln	Leu	Val	Ala	Gly	Asp	Ile	Val	Leu	Asp	Lys	Leu	Gly	Glu	Arg	Leu	
	140					145					150					
gcc	atc	ctc	ctc	aag	gtg	cga	gac	atg	gtc	agc	agc	cat	gtg	gag	cga	711
Ala	Ile	Leu	Leu	Lys	Val	Arg	Asp	Met	Val	Ser	Ser	His	Val	Glu	Arg	
155					160					165				170		
gtg	ttt	cag	atc	tat	gag	caa	cac	gca	gac	aca	gtt	ggc	att	gat	gct	759
Val	Phe	Gln	Ile	Tyr	Glu	Gln	His	Ala	Asp	Thr	Val	Gly	Ile	Asp	Ala	
				175					180					185		
gtc	ctg	cag	cct	tca	gca	gtg	agc	ccc	tct	gtg	gct	gac	atg	ttg	gaa	807
Val	Leu	Gln	Pro	Ser	Ala	Val	Ser	Pro	Ser	Val	Ala	Asp	Met	Leu	Glu	
			190					195					200			
tgg	ttg	cag	gat	att	gag	aga	cat	tat	cga	aag	tcg	tac	ctg	aag	aga	855
Trp	Leu	Gln	Asp	Ile	Glu	Arg	His	Tyr	Arg	Lys	Ser	Tyr	Leu	Lys	Arg	
		205					210					215				
aag	tat	ctt	ctt	tcg	tct	atc	cag	tgg	gga	gac	ttg	gca	aac	ata	caa	903
Lys	Tyr	Leu	Leu	Ser	Ser	Ile	Gln	Trp	Gly	Asp	Leu	Ala	Asn	Ile	Gln	
		220				225					230					
gct	ttg	ccc	aag	gcc	tgg	gac	cga	att	tca	aaa	gac	gaa	cac	cat	gat	951
Ala	Leu	Pro	Lys	Ala	Trp	Asp	Arg	Ile	Ser	Lys	Asp	Glu	His	His	Asp	
235				240					245				250			
ctt	gta	caa	gat	atc	cta	ttg	aat	gtt	tcc	ttc	ttc	ctg	gaa	gag	taa	999
Leu	Val	Gln	Asp	Ile	Leu	Leu	Asn	Val	Ser	Phe	Phe	Leu	Glu	Glu	*	
				255				260					265			
ggaagctatg	cgtatctgga	gaccacgggc	accacagttg	aagactgaca	ttctgaaccc											1059
tgatgtttct	aaagaaacgt	cagtatttca	gtctgacata	tttgaaacat	cagtgccttg											1119
aaccttagga	ctgggtcttg	gggaggatta	gcgcctagat	gtctgatttt	ggagctgcag											1179
catgccaggc	cgtggctgag	agtatgtgag	ccatgccttg	cccttttctg	aggctcaggg											1239
aagtggatgg	agctagagag	gcaacaggaa	agacggtgct	gaagaacata	gtgtctttcc											1299



ctgcctcaga	taaagtgcaa	aggccagtc	gcagggatag	tttcattctc	tccccacgcg	60
atggggctaca	catcacctgc	ccaagctctc	tcccgaacctg	ctagagccta	gagggcgagg	120
gccggagagg	ctgcagccgg	gagtagcacc	gcacatccgg	gaacgccagc	agcgggctga	180
gggctgcata	actgatggaa	ggccgggcgc	ggtaagagcg	tctcggggag	tagggcaagg	240
cggccggggc	cctcccattc	cgccttttct	tcagcgtctc	accgcgggca	ctggctgcga	300
gcgccggggc	acctgcgagt	gtgcgcaggg	actctggaca	cccgcgggcg	cgagctgagg	360
gagcagtc	cacgaggacc	caggcggacc	ctctggcgcc	atg cgc gcc ctc ccc		415
				Met Arg Ala Leu Pro		
				1 5		
ggc ctg ctg gag gcc agg gcg cgt acg ccc cgg ctg ctc ctc ctc cag						463
Gly Leu Leu Glu Ala Arg Ala Arg Thr Pro Arg Leu Leu Leu Leu Gln						
	10		15		20	
tgc ctt ctc gct gcc gcg cgc cca agc tcg gcg gac ggc agt gcc cca						511
Cys Leu Leu Ala Ala Ala Arg Pro Ser Ser Ala Asp Gly Ser Ala Pro						
	25		30		35	
gat tcg cct ttt aca agt cca cct ctc aga gaa gaa ata atg gca aat						559
Asp Ser Pro Phe Thr Ser Pro Pro Leu Arg Glu Glu Ile Met Ala Asn						
	40		45		50	
aac ttt tcc ttg gag agt cat aac ata tca ctg act gaa cat tct agt						607
Asn Phe Ser Leu Glu Ser His Asn Ile Ser Leu Thr Glu His Ser Ser						
	55		60		65	
atg cca gta gaa aaa aat atc act tta gaa agg cct tct aat gta aat						655
Met Pro Val Glu Lys Asn Ile Thr Leu Glu Arg Pro Ser Asn Val Asn						
	70		75		80	85
ctc aca tgc cag ttc aca aca tct ggg gat ttg aat gca gta aat gtg						703
Leu Thr Cys Gln Phe Thr Thr Ser Gly Asp Leu Asn Ala Val Asn Val						
	90		95		100	
act tgg aaa aaa gat ggt gaa caa ctt gag aat aat tat ctt gtc agt						751
Thr Trp Lys Lys Asp Gly Glu Gln Leu Glu Asn Asn Tyr Leu Val Ser						
	105		110		115	
gca aca gga agc acc ttg tat acc caa tac agg ttc acc atc att aat						799
Ala Thr Gly Ser Thr Leu Tyr Thr Gln Tyr Arg Phe Thr Ile Ile Asn						
	120		125		130	
agc aaa caa atg gga agt tat tct tgt ttc ttt cga gag gaa aag gaa						847
Ser Lys Gln Met Gly Ser Tyr Ser Cys Phe Phe Arg Glu Glu Lys Glu						
	135		140		145	
caa agg gga aca ttt aat ttc aaa gtc cct gaa ctt cat ggg aaa aac						895
Gln Arg Gly Thr Phe Asn Phe Lys Val Pro Glu Leu His Gly Lys Asn						
	150		155		160	165
aag cca ttg atc tct tac gta ggg gat tct act gtc ttg aca tgt aaa						943

Lys	Pro	Leu	Ile	Ser	Tyr	Val	Gly	Asp	Ser	Thr	Val	Leu	Thr	Cys	Lys	
				170					175					180		
tgt	caa	aat	tgt	ttt	cct	tta	aat	tgg	acc	tgg	tac	agt	agt	aat	ggg	991
Cys	Gln	Asn	Cys	Phe	Pro	Leu	Asn	Trp	Thr	Trp	Tyr	Ser	Ser	Asn	Gly	
			185					190					195			
agt	gta	aag	gtt	cct	gtt	ggg	gtt	caa	atg	aat	aaa	tat	gtg	atc	aat	1039
Ser	Val	Lys	Val	Pro	Val	Gly	Val	Gln	Met	Asn	Lys	Tyr	Val	Ile	Asn	
		200					205					210				
gga	aca	tat	gct	aac	gaa	aca	aag	ctg	aag	ata	aca	caa	ctt	ttg	gag	1087
Gly	Thr	Tyr	Ala	Asn	Glu	Thr	Lys	Leu	Lys	Ile	Thr	Gln	Leu	Leu	Glu	
	215					220					225					
gaa	gat	ggg	gaa	tct	tac	tgg	tgc	cgt	gca	cta	ttc	caa	tta	ggc	gag	1135
Glu	Asp	Gly	Glu	Ser	Tyr	Trp	Cys	Arg	Ala	Leu	Phe	Gln	Leu	Gly	Glu	
230				235				240						245		
agt	gaa	gaa	cac	att	gag	ctt	gtg	gtg	ctg	agc	tat	ttg	gtg	ccc	ctc	1183
Ser	Glu	Glu	His	Ile	Glu	Leu	Val	Val	Leu	Ser	Tyr	Leu	Val	Pro	Leu	
			250					255						260		
aaa	cca	ttt	ctt	gta	ata	gtg	gct	gag	gtg	att	ctt	tta	gtg	gcc	acc	1231
Lys	Pro	Phe	Leu	Val	Ile	Val	Ala	Glu	Val	Ile	Leu	Leu	Val	Ala	Thr	
			265					270					275			
att	ctg	ctt	tgt	gaa	aag	tac	aca	caa	aag	aaa	aag	aag	cac	tca	gat	1279
Ile	Leu	Leu	Cys	Glu	Lys	Tyr	Thr	Gln	Lys	Lys	Lys	Lys	His	Ser	Asp	
	280						285					290				
gag	ggg	aaa	gaa	ttt	gag	cag	att	gaa	cag	ctg	aaa	tca	gat	gat	agc	1327
Glu	Gly	Lys	Glu	Phe	Glu	Gln	Ile	Glu	Gln	Leu	Lys	Ser	Asp	Asp	Ser	
	295					300				305						
aat	ggg	ata	gaa	aat	aat	gtc	ccc	agg	cat	aga	aaa	aat	gag	tct	ctg	1375
Asn	Gly	Ile	Glu	Asn	Asn	Val	Pro	Arg	His	Arg	Lys	Asn	Glu	Ser	Leu	
310				315				320					325			
ggc	cag	tga	atacaaaa	acatcatgtc	gagaatcatt	ggaagatata	cagagttcgt									1431
Gly	Gln	*														
atttcagctt	tgtttatacct	tctgttaag	agcctctgag	tttttagttt	taaaaggatg											1491
aaaagcttat	gcaacatgct	cagcaggagc	ttcatcaacg	atatatgtca	gatctaaagg											1551
tatattttca	ttctgtaatt	atgttacata	aaagcaatgt	aatcagaat	aatatgtta											1611
gaccagaata	aaattaatta	tattctgggc	ttcaaaggac	acacagaaca	gatatcagca											1671
gaatcactta	atacttcata	gaacaaaaat	cactcaaaac	ctgtttataa	ccaaagaatt											1731
catgaaaaag	aaagcctttg	ccatttgtct	tagaaagtta	tttttttaaa	aaaaa											1786

<210> 26  
 <211> 1412  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> (107)..(1189)

<400> 26

taagcttgcg gccgccgctc aggtcaggca gccccagcgt ccccgggccc tcggtccac 60

cgagtgccgg ctcccgcgct ctgcgcaagc ccttggaatt ttcaaa atg aat tat 115  
 Met Asn Tyr  
 1

aca gag tcc agc cca ttg aga gaa tca act gcc ata ggt ttt aca cct 163  
 Thr Glu Ser Ser Pro Leu Arg Glu Ser Thr Ala Ile Gly Phe Thr Pro  
 5 10 15

gag tta gaa agt atc ata cct gtg cct tcc aat aag acc act tgt gaa 211  
 Glu Leu Glu Ser Ile Ile Pro Val Pro Ser Asn Lys Thr Thr Cys Glu  
 20 25 30 35

aac tgg aga gag ata cat cat ctg gtt ttt cat gta gca aat att tgt 259  
 Asn Trp Arg Glu Ile His His Leu Val Phe His Val Ala Asn Ile Cys  
 40 45 50

ttt gca gtt ggg ttg gtt att cca act act ctt cac ctt cat atg ata 307  
 Phe Ala Val Gly Leu Val Ile Pro Thr Thr Leu His Leu His Met Ile  
 55 60 65

ttt ctt agg gga atg tta act cta gga tgt acc ctt tat atc gtc tgg 355  
 Phe Leu Arg Gly Met Leu Thr Leu Gly Cys Thr Leu Tyr Ile Val Trp  
 70 75 80

gcc act ctc tac cga tgt gcc ttg gat ata atg atc tgg aac tct gtg 403  
 Ala Thr Leu Tyr Arg Cys Ala Leu Asp Ile Met Ile Trp Asn Ser Val  
 85 90 95

ttc ttg ggt gtc aac att ttg cat ctg tcg tat ctt tta tac aag aag 451  
 Phe Leu Gly Val Asn Ile Leu His Leu Ser Tyr Leu Leu Tyr Lys Lys  
 100 105 110 115

aga ccg gta aag att gaa aag gaa ctc agt ggc atg tac cgg cga ttg 499  
 Arg Pro Val Lys Ile Glu Lys Glu Leu Ser Gly Met Tyr Arg Arg Leu  
 120 125 130

ttt gaa cca ctc cgt gtg cct cca gat ttg ttc aga aga cta act gga 547  
 Phe Glu Pro Leu Arg Val Pro Pro Asp Leu Phe Arg Arg Leu Thr Gly  
 135 140 145

cag ttt tgc atg atc caa acc ttg aaa aag ggc caa act tat gct gca 595  
 Gln Phe Cys Met Ile Gln Thr Leu Lys Lys Gly Gln Thr Tyr Ala Ala  
 150 155 160

GenBank# F04936

gag gat aaa acc tca gtt gat gac cgt ctg agt att ctc ttg aag gga Glu Asp Lys Thr Ser Val Asp Asp Arg Leu Ser Ile Leu Leu Lys Gly 165 170 175	643
aaa atg aag gtc tcc tat cga gga cat ttt ctg cat aac att tac ccc Lys Met Lys Val Ser Tyr Arg Gly His Phe Leu His Asn Ile Tyr Pro 180 185 190 195	691
tgt gcc ttt ata gat tct cct gaa ttt aga tca act cag atg cac aaa Cys Ala Phe Ile Asp Ser Pro Glu Phe Arg Ser Thr Gln Met His Lys 200 205 210	739
ggg gaa aaa ttc cag gtc acc att att gca gat gat aac tgc aga ttt Gly Glu Lys Phe Gln Val Thr Ile Ile Ala Asp Asp Asn Cys Arg Phe 215 220 225	787
tta tgc tgg tca aga gaa aga tta aca tac ttt ctg gaa tca gaa cct Leu Cys Trp Ser Arg Glu Arg Leu Thr Tyr Phe Leu Glu Ser Glu Pro 230 235 240	835
ttc ttg tat gaa atc ttt agg tat ctt att gga aaa gac atc aca aat Phe Leu Tyr Glu Ile Phe Arg Tyr Leu Ile Gly Lys Asp Ile Thr Asn 245 250 255	883
aag ctc tac tca ttg aat gat ccc acc tta aat gat aaa aaa gcc aaa Lys Leu Tyr Ser Leu Asn Asp Pro Thr Leu Asn Asp Lys Lys Ala Lys 260 265 270 275	931
aag ctg gaa cat cag ctc agc ctc tgc aca cag atc tcc atg ttg gaa Lys Leu Glu His Gln Leu Ser Leu Cys Thr Gln Ile Ser Met Leu Glu 280 285 290	979
atg agg aac agt ata gcc agc tcc agt gac agt gac gac ggc ttg cac Met Arg Asn Ser Ile Ala Ser Ser Ser Asp Ser Asp Asp Gly Leu His 295 300 305	1027
cag ttt ctt cgg ggt acc tcc agc atg tcc tct ctt cat gtg tca tcc Gln Phe Leu Arg Gly Thr Ser Ser Met Ser Ser Leu His Val Ser Ser 310 315 320	1075
cca cac cag cga gcc tct gcc aag atg aaa ccg ata gaa gaa gga gca Pro His Gln Arg Ala Ser Ala Lys Met Lys Pro Ile Glu Glu Gly Ala 325 330 335	1123
gaa gat gat gat gac gtt ttt gaa ccg gca tct cca aat aca ttg aaa Glu Asp Asp Asp Val Phe Glu Pro Ala Ser Pro Asn Thr Leu Lys 340 345 350 355	1171
gtc cat cag ctg cct tga tcagag agagaattca ggttaccaag acggaaggtg Val His Gln Leu Pro * 360	1225
tcttgaagag atcctgaaaa ataccagcac tttttcatgg ctttttaggtt attctgcttt	1285
agtgcattcca gactggtaga gtcggaggga ggaagtgagg aagggtcaag gatggaagag	1345
ttcttttact tacccttttt attagtcagc ttttaaagta attgtttttac tgaaaaaaaa	1405



aaaaaaa

1412

<210> 27  
<211> 1861  
<212> DNA  
<213> Homo sapiens

<220>  
<221> CDS  
<222> (325) .. (1242)

<400> 27

tgtggtaccg gtccggaatt cccgggtcga cgatttcgtc cggcgagaga ggacccaggt	60
gtcctggcag tgggcgccgc ggggcacacg ctgggccaaag gtgcaggcgg ccaggggtggg	120
agactgttcg ccccgccctg agtactccta tcttgtttct ccacctgttc gggagttgga	180
gatgtgcacc taaaggaggc gcatctgggg acggacacat ctggcactga ggccctcgcc	240
acctgcctcg ccacctggcg accctgaccc caccacactg ccttgagagt cgctcaaaag	300
tagggcccca gggctcgcag cagc atg ggc acc gag aaa gaa agc cca gag	351
Met Gly Thr Glu Lys Glu Ser Pro Glu	
1 5	
ccc gac tgc cag aaa cag ttc cag gct gca gtg agc gtc atc cag aac	399
Pro Asp Cys Gln Lys Gln Phe Gln Ala Ala Val Ser Val Ile Gln Asn	
10 15 20 25	
ctg ccc aag aac ggt tct tac cgc ccc tcc tat gaa gag atg ctg cga	447
Leu Pro Lys Asn Gly Ser Tyr Arg Pro Ser Tyr Glu Glu Met Leu Arg	
30 35 40	
ttc tac agt tac tac aag cag gcc acc atg ggg ccc tgc ctg gtc ccc	495
Phe Tyr Ser Tyr Tyr Lys Gln Ala Thr Met Gly Pro Cys Leu Val Pro	
45 50 55	
cgg ccc ggg ttc tgg gac ccc att gga cga tat aag tgg gac gcc tgg	543
Arg Pro Gly Phe Trp Asp Pro Ile Gly Arg Tyr Lys Trp Asp Ala Trp	
60 65 70	
aac agt ctg ggc aag atg agc agg gag gag gcc atg tct gcc tac atc	591
Asn Ser Leu Gly Lys Met Ser Arg Glu Glu Ala Met Ser Ala Tyr Ile	
75 80 85	
act gaa atg aaa ctg gtg gca cag aag gtg atc gac aca gtg ccc ctg	639
Thr Glu Met Lys Leu Val Ala Gln Lys Val Ile Asp Thr Val Pro Leu	
90 95 100 105	
ggg gag gtg gca gag gac atg ttt ggt tac ttc gag ccc ctg tac cag	687
Gly Glu Val Ala Glu Asp Met Phe Gly Tyr Phe Glu Pro Leu Tyr Gln	
110 115 120	

gtg atc cct gac atg ccg agg ccc cca gag acc ttc ctg aga agg gtc Val Ile Pro Asp Met Pro Arg Pro Pro Glu Thr Phe Leu Arg Arg Val 125 130 135	735
aca ggt tgg aaa gag cag gtt gtg aat gga gat gtt ggg gct gtt tca Thr Gly Trp Lys Glu Gln Val Val Asn Gly Asp Val Gly Ala Val Ser 140 145 150	783
gag cct ccc tgc ctc ccc aag gaa ccg gca ccc cca agc cca gag tcc Glu Pro Pro Cys Leu Pro Lys Glu Pro Ala Pro Pro Ser Pro Glu Ser 155 160 165	831
cat tca ccc agg gac ctg gac tcc gag gtt ttc tgt gat tcc ctg gag His Ser Pro Arg Asp Leu Asp Ser Glu Val Phe Cys Asp Ser Leu Glu 170 175 180 185	879
cag ctg gag cct gag ctg gtt tgg aca gag cag cgg gca gca tct gga Gln Leu Glu Pro Glu Leu Val Trp Thr Glu Gln Arg Ala Ala Ser Gly 190 195 200	927
gga aag cgt gat ccc agg aac agc ccc gtg ccc ccc aca aag aaa gag Gly Lys Arg Asp Pro Arg Asn Ser Pro Val Pro Pro Thr Lys Lys Glu 205 210 215	975
ggg ttg cgg ggc agc ccg ccg ggg ccc cag gag ttg gac gtg tgg ctg Gly Leu Arg Gly Ser Pro Pro Gly Pro Gln Glu Leu Asp Val Trp Leu 220 225 230	1023
ctg ggg aca gtt cga gca cta cag gag agc atg cag gag gtg cag gcg Leu Gly Thr Val Arg Ala Leu Gln Glu Ser Met Gln Glu Val Gln Ala 235 240 245	1071
agg gtg cag agc ctg gag agc atg ccc cgg ccc cct gag cag agg ccg Arg Val Gln Ser Leu Glu Ser Met Pro Arg Pro Pro Glu Gln Arg Pro 250 255 260 265	1119
cag ccc agg ccc agt gct cgg cca tgg ccc ctt ggg ctc ccg ggg ccc Gln Pro Arg Pro Ser Ala Arg Pro Trp Pro Leu Gly Leu Pro Gly Pro 270 275 280	1167
gcg ctg ctc ttc ttc ctc ctg tgg ccc ttc gtc gtc cag tgg ctc ttc Ala Leu Leu Phe Phe Leu Leu Trp Pro Phe Val Val Gln Trp Leu Phe 285 290 295	1215
cga atg ttt cgg acc caa aag agg tga ctgtc agtggagggg tctctgcagc Arg Met Phe Arg Thr Gln Lys Arg *	1267
160 300 305	
caactgagac tatcttgctg tgccctgagc ctctcctaggg tttagaagaa cagcattcaa	1327
aattccccgt cctgtcagtg ttgtccttcg cacctcctcc cctaaagcag cgcggggggc	1387
aaataagacc ccacccctcc ctgcagcttc acagggagcg ttccttcct ccccgcaacc	1447
accccaggct cccctgggag gctgcagttg tggtacacgt ccccggtgct gggttggccg	1507

```

tgactcgggg gcggggcgat cgggtctcag cccctgcctt cccagtcctc tgggtcaccc 1567
gaattttccc acccctgctt ctccccgagg aggttgagct cttgagcaag ttgggacttg 1627
ggccgggggcc tggaagaatg attggctggg aggccgcggg agggaggcca ggaggcccgg 1687
accagtggg aggagtgagc agggccccggg ggaggggggat gagcgcagtt tgctcgcttt 1747
cctccccctgc cggccccctc cgccccaca cacactcggg acgtcttcat tgaagattca 1807
cttacaaagg aatgtttcac taaataaaag aaaaccagaa aaaaaaaaaa aaaa 1861

```

```

<210> 28
<211> 997
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> CDS
<222> (151)..(912)

```

```

<400> 28
taagcttgcg gccgccccga cgcttcggtt cggtcgctcg gcacgtcggc cctgcctccc 60
ggaccaccct gcggggcgca ccagcgatct ggggtgcggg gctcggcctc cctgcgctcc 120
tggtgacgg tgtgaccttg ggtcttaacc      atg aac ttc tct gga gga ggg 171
                                   Met Asn Phe Ser Gly Gly Gly
                                   1           5

agg cag gaa gca gca ggg tcc agg ggt aga agg gct ccc aga ccc cga 219
Arg Gln Glu Ala Ala Gly Ser Arg Gly Arg Arg Ala Pro Arg Pro Arg
      10           15           20

gaa cag gac cga gac gtg cag ctg tcc aag gct ctg tcc tat gcc ctg 267
Glu Gln Asp Arg Asp Val Gln Leu Ser Lys Ala Leu Ser Tyr Ala Leu
      25           30           35

cgc cat ggg gcc ttg aag ctg ggg ctt ccc atg gga gct gat ggc ttc 315
Arg His Gly Ala Leu Lys Leu Gly Leu Pro Met Gly Ala Asp Gly Phe
      40           45           50           55

gtg ccc ctg ggc acc ctc ctg cag ttg ccc cag ttc cgc ggc ttc tct 363
Val Pro Leu Gly Thr Leu Leu Gln Leu Pro Gln Phe Arg Gly Phe Ser
      60           65           70

gct gaa gat gtg cag cgc gtg gtg gac acc aat agg aag cag cgg ttc 411
Ala Glu Asp Val Gln Arg Val Val Asp Thr Asn Arg Lys Gln Arg Phe
      75           80           85

gcc ctg cag ctg ggg gat ccc agc act ggc ctt ctc atc cgg gcc aac 459
Ala Leu Gln Leu Gly Asp Pro Ser Thr Gly Leu Leu Ile Arg Ala Asn
      90           95           100

```

cag	ggc	cat	tcc	ctg	cag	gta	cct	aag	ttg	gag	ctg	atg	ccc	ctg	gag	507
Gln	Gly	His	Ser	Leu	Gln	Val	Pro	Lys	Leu	Glu	Leu	Met	Pro	Leu	Glu	
105						110			115							
aca	ccg	cag	gcc	ctg	ccc	ccg	atg	cta	gtc	cat	ggg	aca	ttc	tgg	aag	555
Thr	Pro	Gln	Ala	Leu	Pro	Pro	Met	Leu	Val	His	Gly	Thr	Phe	Trp	Lys	
120			125						130			135				
cac	tgg	cca	tcc	atc	cta	ctc	aaa	ggc	ctg	tcc	tgc	cag	gga	agg	acg	603
His	Trp	Pro	Ser	Ile	Leu	Leu	Lys	Gly	Leu	Ser	Cys	Gln	Gly	Arg	Thr	
			140						145			150				
cac	att	cac	ctg	gcc	cca	gga	ctg	cct	gga	gac	ccc	ggg	atc	atc	agt	651
His	Ile	His	Leu	Ala	Pro	Gly	Leu	Pro	Gly	Asp	Pro	Gly	Ile	Ile	Ser	
155						160						165				
ggc	atg	cgg	tcc	cat	tgt	gaa	ata	gct	gtg	ttc	atc	gat	gga	ccc	ctg	699
Gly	Met	Arg	Ser	His	Cys	Glu	Ile	Ala	Val	Phe	Ile	Asp	Gly	Pro	Leu	
170						175			180							
gct	ctg	gca	gat	gga	ata	ccc	ttc	ttc	cgc	tct	gcc	aat	ggg	gtg	att	747
Ala	Leu	Ala	Asp	Gly	Ile	Pro	Phe	Phe	Arg	Ser	Ala	Asn	Gly	Val	Ile	
185						190			195							
ctg	act	cca	ggg	aat	act	gat	ggc	ttc	ctc	ctt	ccc	aag	tac	ttc	aag	795
Leu	Thr	Pro	Gly	Asn	Thr	Asp	Gly	Phe	Leu	Leu	Pro	Lys	Tyr	Phe	Lys	
200			205						210			215				
gag	gcc	ctg	cag	cta	cgc	cct	acc	cga	aag	ccc	ctt	tcc	ttg	gct	ggg	843
Glu	Ala	Leu	Gln	Leu	Arg	Pro	Thr	Arg	Lys	Pro	Leu	Ser	Leu	Ala	Gly	
			220						225			230				
gat	gaa	gag	aca	gag	tgt	cag	agt	agc	ccc	aag	cac	agc	tcc	aga	gaa	891
Asp	Glu	Glu	Thr	Glu	Cys	Gln	Ser	Ser	Pro	Lys	His	Ser	Ser	Arg	Glu	
			235			240			245							
agg	agg	agg	atc	caa	caa	taa	aa	tattaattta	taaaaaaagaa	atttttaaaaa						944
Arg	Arg	Arg	Ile	Gln	Gln	*										
250																
gtaacaagaa agaactcgtt tqaaaccatg tttcatcaaaa aaaaaaaaaa aaa																997

```
<210> 29
<211> 1138
<212> DNA
<213> Homo sapiens
```

<400> 29  
ttagataccc cttagatatg ccgttacgcc tgcggtagcc gaccggaatt cccgggtcga 60

cccacgcgtc	cggttgagaa	ggctacatgt	gacgagaatg	attctgtttc	aaatatagcc		120
acagaaataa	aggaggggaca	acaatctggg	acagtgtctc	ctcagaaaca	atcgggcccag		180
aaggttatat	ttaaaaagaa	agtttctctt	ttgaatatgt	ccacaagaat	a atg ggc	237	
					Met Gly		
					1		
ggt ggg aaa tct gga aca gtg tct tct cag aaa caa cca gcc tca aag						285	
Gly Gly Lys 5 Ser Gly Thr Val Ser Ser Gln Lys Gln Pro Ala Ser Lys							
act gca agt gac aag aca gat tct gct ttg aat aca gct aca gaa ata						333	
Thr Ala Ser Asp Lys Thr Asp Ser Ala Leu Asn Thr Ala Thr Glu Ile							
aag gat gga cta caa tgt ggg aca gtg tct tct cag aaa caa caa gcc						381	
Lys Asp Gly Leu Gln Cys Gly Thr Val Ser Ser Gln Lys Gln Gln Ala							
ttg aag gct aca act gac gag gaa ggt tct gtt tct aat ata gcc aca						429	
Leu Lys Ala Thr Thr Asp Glu Glu Gly Ser Val Ser Asn Ile Ala Thr							
gaa ata aag gat gga gaa aaa tct ggg aca gtg tct tct cag aaa aaa						477	
Glu Ile Lys Asp Gly Glu Lys Ser Gly Thr Val Ser Ser Gln Lys Lys							
cca gcc ttg aag gcc aca agt gat gag aaa gat tct ttt tcg aat ata						525	
Pro Ala Leu Lys Ala Thr Ser Asp Glu Lys Asp Ser Phe Ser Asn Ile							
acc aga gaa aaa aag gat gga gaa ata tct agg aca gtg tct tct cag						573	
Thr Arg Glu Lys Lys Asp Gly Glu Ile Ser Arg Thr Val Ser Ser Gln							
aaa cca cca gcc ttg aag gct aca agt gtc aag gaa gat tct gtt ttg						621	
Lys Pro Pro Ala Leu Lys Ala Thr Ser Val Lys Glu Asp Ser Val Leu							
aat ata gcc aga gaa aaa aag gat gga gaa aaa tct agg aca gtg tct						669	
Asn Ile Ala Arg Glu Lys Lys Asp Gly Glu Lys Ser Arg Thr Val Ser							
ttt gac caa cca cca ggc ttg aag gct aca aga gac gag aaa gat tct						717	
Phe Asp Gln Pro Pro Gly Leu Lys Ala Thr Arg Asp Glu Lys Asp Ser							
ctt ttg aat ata gcc aga gga aaa gag gat gga gaa aaa act agg aga						765	
Leu Leu Asn Ile Ala Arg Gly Lys Glu Asp Gly Glu Lys Thr Arg Arg							
gtg tct tct cgg aaa aaa cca gcc ttg aag gct aca agt gat gag aaa						813	
Val Ser Ser Arg Lys Lys Pro Ala Leu Lys Ala Thr Ser Asp Glu Lys							
gat tct ttt tcg aat ata acc aqa qaa aaa aaq qat qqa qaa aca tct						861	

```

Asp Ser Phe Ser Asn Ile Thr Arg Glu Lys Lys Asp Gly Glu Thr Ser
195                200                205                210

agg aca gtg tct tct cag aaa cca cca gcc ttg aag gct aca agt gac      909
Arg Thr Val Ser Ser Gln Lys Pro Pro Ala Leu Lys Ala Thr Ser Asp
                215                220                225

gag gaa gat tct gtt ttg agt ata gcc aga gaa gaa aag gat gga gaa      957
Glu Glu Asp Ser Val Leu Ser Ile Ala Arg Glu Glu Lys Asp Gly Glu
                230                235                240

aaa tct agg aca gtg tct tct gag caa cca cca ggc ttg aag tgt ctt      1005
Lys Ser Arg Thr Val Ser Ser Glu Gln Pro Pro Gly Leu Lys Cys Leu
                245                250                255

ctc gga aaa aag cag cct tga ag gctacaagtg atgagaaaga ttcttttttca      1058
Leu Gly Lys Lys Gln Pro *
                260                265

aatataacca gagaaagaaa ggatggagaa acatctagga cagtgtcttc tcagaaacca      1118

ccagccttga aggctacaag                                          1138

<210> 30
<211> 1270
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (274)..(1173)

<400> 30
tcatcacccg tagcatccgg ttaagaggct ctccacctcg ctcatTTTtgg totaaactgga      60

tttacacaaa ctcaactgtgc acgtgcctgg ggcgggcaaa cccggaaacg acgccgagtc      120

ccgcgggatc caggaggagc caactgcgcc ggaggagggg tttcggccga cacgtcggga      180

ttggcggctg cagccagggg tcttccgacg ctgggcttcc gtgagcggcg ctctgccaga      240

tctctggacc ggattcgtcc cattctcgtc ctc atg gtg gac aag aaa ctg gtg      294
                        Met Val Asp Lys Lys Leu Val
                        1                        5

gtg gtt ttc gga ggc aca ggt gcc cag ggt ggc tcc gtg gcc cgc aca      342
Val Val Phe Gly Gly Thr Gly Ala Gln Gly Gly Ser Val Ala Arg Thr
                10                15                20

ctc ctg gaa gat ggg aca ttc aag gtt cga gtg gtg acc cga aac cct      390
Leu Leu Glu Asp Gly Thr Phe Lys Val Arg Val Thr Arg Asn Pro
                25                30                35

agg aag aag gca gca aag gag ctg agg ctg caa ggt gca gaa gta gtg      438

```

Arg 40	Lys	Lys	Ala	Ala	Lys 45	Glu	Leu	Arg	Leu	Gln	Gly	Ala	Glu	Val	Val 55	
cag Gln	gga Gly	gac Asp	caa Gln	gat Asp 60	gac Asp	cag Gln	gtc Val	atc Ile	atg Met 65	gag Glu	ctg Leu	gcc Ala	ctg Leu	aat Asn 70	ggg Gly	486
gct Ala	tac Tyr	gcc Ala	acc Thr 75	ttc Phe	atc Ile	gtg Val	acc Thr	aat Asn 80	tac Tyr	tgg Trp	gag Glu	agc Ser	tgc Cys 85	agc Ser	cag Gln	534
gag Glu	cag Gln	gag Glu 90	gtc Val	aag Lys	cag Gln	ggg Gly	aag Lys 95	ctg Leu	ctc Leu	gct Ala	gat Asp	ctg Leu 100	gcc Ala	agg Arg	cgc Arg	582
ctg Leu 105	ggc Gly	ctc Leu	cac His	tat Tyr	gtg Val 110	gtc Val	tac Tyr	agc Ser	ggc Gly	ctg Leu 115	gag Glu	aac Asn	atc Ile	aag Lys	aag Lys	630
ctg Leu 120	acg Thr	gca Ala	ggg Gly	aga Arg	ttg Leu 125	gcc Ala	gcc Ala	gcg Ala	cac His	ttt Phe 130	gac Asp	ggc Gly	aaa Lys	ggg Gly	gag Glu 135	678
gtg Val	gag Glu	gaa Glu	tat Tyr	ttc Phe 140	cgg Arg	gac Asp	att Ile	ggc Gly	gtt Val 145	ccc Pro	atg Met	acc Thr	agt Ser	gtg Val 150	cgg Arg	726
ctg Leu	ccc Pro	tgc Cys	tat Tyr 155	ttt Phe	gag Glu	aac Asn	ctc Leu 160	ctc Leu	tcc Ser	cac His	ttc Phe	ttg Leu 165	ccc Pro	cag Gln	aaa Lys	774
gcc Ala	cca Pro	gac Asp 170	gga Gly	aag Lys	agc Ser	tac Tyr	ttg Leu 175	ctg Leu	agc Ser	ttg Leu	ccc Pro	aca Thr 180	ggt Gly	gac Asp	gtt Val	822
ccc Pro 185	atg Met	gat Asp	ggc Gly	atg Met	tcc Ser	gtg Val 190	tct Ser	gac Asp	ctg Leu	ggt Gly	cct Pro	gtg Val	gtg Val	ctc Leu	agc Ser	870
ctt Leu 200	ttg Leu	aag Lys	atg Met	cca Pro 205	gaa Glu	aaa Lys	tac Tyr	gtc Val	ggc Gly	cag Gln 210	aac Asn	atc Ile	ggg Gly	ctg Leu	agc Ser 215	918
act Thr	tgc Cys	agg Arg	cac His 220	acg Thr	gcc Ala	gag Glu	gag Glu	tac Tyr	gct Ala 225	gcc Ala	ctg Leu	ctc Leu	acc Thr	aag Lys 230	cac His	966
acc Thr	cgc Arg	aag Lys	gtc Val 235	gtg Val	cac His	gat Asp	gcc Ala	aag Lys 240	atg Met	act Thr	cct Pro	gag Glu	gac Asp 245	tac Tyr	gaa Glu	1014
aag Lys	ctt Leu	ggc Gly 250	ttt Phe	ccc Pro	ggt Gly	gcc Ala	cgg Arg 255	gac Asp	ctg Leu	gcc Ala	aac Asn	atg Met 260	ttc Phe	cgt Arg	ttc Phe	1062
tat Tyr	gcc Ala	ctg Leu	aga Arg	ccc Pro	gac Asp	cgt Arg	gac Asp	atc Ile	gag Glu	ctg Leu	acc Thr	ctg Leu	aga Arg	ctc Leu	aac Asn	1110

```

265                270                275
ccc aag gcc ctg acg ctg gac cag tgg ctg gaa cag cac aaa ggg gac 1158
Pro Lys Ala Leu Thr Leu Asp Gln Trp Leu Glu Gln His Lys Gly Asp
280                285                290                295

ttc aac ctg ctg tga cctgccccgcc tcgcggcccc ttgtggggat cgggggcacc 1213
Phe Asn Leu Leu *
300

agaggggagcag aggcaccaac atctgaataa agccattgtt ctcccaaaaa aaaaaaa 1270

```

```

<210> 31
<211> 3053
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (454)..(2802)

<220>
<221> misc_feature
<222> (1)...(3053)
<223> n = a,t,c or g

```

```

<400> 31
gggaaacgga caacattcgt tttgagcaaa agaagtagag tatatctgca atactgaaga 60
aaaaaatggt ctttaaaact atgaagtggg taatgggatt aaagtattca aggaatgaga 120
tggaatcctt ggtaatgggt agaaacatag gtaaattctc ctttccact taccaccaac 180
ctgtgtgaac aacaatcatt tctcattaat agtcaagaat gtgaagagat catgagtana 240
aagtcttgat cagaaattgt gtatagctgt gtccagcata aaacottata taaaatgcat 300
attcatagaa gatgctcaag tatgtcatga aaaatatgaa gtttatttaa atatttatcc 360
aaattaaaga tttnctgtat tctctaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaagcct 420
cgtgccgaat tcggcacgag ggctgtcgcc acc atg gct ccg cac cgc ccc gcg 474
                        Met Ala Pro His Arg Pro Ala
                        1                5

ccc gcg ctg ctt tgc gcg ctg tcc ctg gcg ctg tgc gcg ctg tcg ctg 522
Pro Ala Leu Leu Cys Ala Leu Ser Leu Ala Leu Cys Ala Leu Ser Leu
10                15                20

ccc gtc cgc gcg gcc act gcg tcg cgg ggg gcg tcc cag gcg ggg gcg 570
Pro Val Arg Ala Ala Thr Ala Ser Arg Gly Ala Ser Gln Ala Gly Ala
25                30                35

ccc cag ggg cgg gtg ccc gag gcg cgg ccc aac agc atg gtg gtg gaa 618

```



Pro 40	Gln	Gly	Arg	Val	Pro 45	Glu	Ala	Arg	Pro	Asn 50	Ser	Met	Val	Val	Glu 55	
cac His	ccc Pro	gag Glu	ttc Phe	ctc Leu 60	aag Lys	gca Ala	ggg Gly	aag Lys	gag Glu 65	cct Pro	ggc Gly	ctg Leu	cag Gln 70	atc Ile	tgg Trp	666
cgt Arg	gtg Val	gag Glu	aag Lys 75	ttc Phe	gat Asp	ctg Leu	gtg Val	ccc Pro 80	gtg Val	ccc Pro	acc Thr	aac Asn	ctt Leu 85	tat Tyr	gga Gly	714
gac Asp	ttc Phe	ttc Phe 90	acg Thr	ggc Gly	gac Asp	gcc Ala	tac Tyr 95	gtc Val	atc Ile	ctg Leu	aag Lys 100	aca Thr	gtg Val	cag Gln	ctg Leu	762
agg Arg 105	aac Asn	gga Gly	aat Asn	ctg Leu	cag Gln	tat Tyr 110	gac Asp	ctc Leu	cac His	tac Tyr	tgg Trp 115	ctg Leu	ggc Gly	aat Asn	gag Glu	810
tgc Cys 120	agc Ser	cag Gln	gat Asp	gag Glu	agc Ser 125	ggg Gly	gcg Ala	gcc Ala	gcc Ala	atc Ile 130	ttt Phe	acc Thr	gtg Val	cag Gln	ctg Leu 135	858
gat Asp	gac Asp	tac Tyr	ctg Leu	aac Asn 140	ggc Gly	cgg Arg	gcc Ala	gtg Val	cag Gln 145	cac His	cgt Arg	gag Glu	gtc Val	cag Gln 150	ggc Gly	906
ttc Phe	gag Glu	tcg Ser 155	gcc Ala	acc Thr	ttc Phe	cta Leu	ggc Gly 160	tac Tyr	ttc Phe	aag Lys	tct Ser	ggc Gly 165	ctg Leu 165	aag Lys	tac Tyr	954
aag Lys 170	aaa Lys	gga Gly	ggt Gly	gtg Val	gca Ala	tca Ser	gga Gly 175	ttc Phe	aag Lys	cac His	gtg Val 180	gta Val	ccc Pro	aac Asn	gag Glu	1002
gtg Val 185	gtg Val	gtg Val	cag Gln	aga Arg	ctc Leu	ttc Phe 190	cag Gln	gtc Val	aaa Lys	ggg Gly 195	cgg Arg	cgt Arg	gtg Val	gtc Val	cgt Arg	1050
gcc Ala 200	acc Thr	gag Glu	gta Val	cct Pro	gtg Val 205	tcc Ser	tgg Trp	gag Glu	agc Ser 210	ttc Phe	aac Asn	aat Asn	ggc Gly	gac Asp	tgc Cys 215	1098
ttc Phe	atc Ile	ctg Leu	gac Asp 220	ctg Leu	ggc Gly	aac Asn	aac Asn	atc Ile 225	cac His	cag Gln	tgg Trp	tgt Cys	ggt Gly 230	tcc Ser	aac Asn	1146
agc Ser	aat Asn	cgg Arg	tat Tyr 235	gaa Glu	aga Arg	ctg Leu	aag Lys	gcc Ala 240	aca Thr	cag Gln	gtg Val	tcc Ser	aag Lys 245	ggc Gly	atc Ile	1194
cgg Arg	gac Asp	aac Asn 250	gag Glu	cgg Arg	agt Ser	ggc Gly	cgg Arg 255	gcc Ala	cga Arg	gtg Val	cac His 260	gtg Val	tct Ser	gag Glu	gag Glu	1242
ggc Gly	act Thr	gag Glu	ccc Pro	gag Glu	gcg Ala	atg Met	ctc Leu	cag Gln	gtg Val	ctg Leu	ggc Gly	ccc Pro	aag Lys	ccg Pro	gct Ala	1290



gct gca tct gcc atc ctg act gct cag ctg gat gag gag ctg gga ggt	2010
Ala Ala Ser Ala Ile Leu Thr Ala Gln Leu Asp Glu Glu Leu Gly Gly	
505 510 515	
acc cct gtc cag agc cgt gtg gtc caa ggc aag gag ccc gcc cac ctc	2058
Thr Pro Val Gln Ser Arg Val Val Gln Gly Lys Glu Pro Ala His Leu	
520 525 530 535	
atg agc ctg ttt ggt ggg aag ccc atg atc atc tac aag ggc ggc acc	2106
Met Ser Leu Phe Gly Gly Lys Pro Met Ile Ile Tyr Lys Gly Gly Thr	
540 545 550	
tcc cgc gag ggc ggg cag aca gcc cct gcc agc acc cgc ctc ttc cag	2154
Ser Arg Glu Gly Gly Gln Thr Ala Pro Ala Ser Thr Arg Leu Phe Gln	
555 560 565	
gtc cgc gcc aac agc gct gga gcc acc cgg gct gtt gag gta ttg cct	2202
Val Arg Ala Asn Ser Ala Gly Ala Thr Arg Ala Val Glu Val Leu Pro	
570 575 580	
aag gct ggt gca ctg aac tcc aac gat gcc ttt gtt ctg aaa acc ccc	2250
Lys Ala Gly Ala Leu Asn Ser Asn Asp Ala Phe Val Leu Lys Thr Pro	
585 590 595	
tca gcc gcc tac ctg tgg gtg ggt aca gga gcc agc gag gca gag aag	2298
Ser Ala Ala Tyr Leu Trp Val Gly Thr Gly Ala Ser Glu Ala Glu Lys	
600 605 610 615	
acg ggg gcc cag gag ctg ctc agg gtg ctg cgg gcc caa cct gtg cag	2346
Thr Gly Ala Gln Glu Leu Leu Arg Val Leu Arg Ala Gln Pro Val Gln	
620 625 630	
gtg gca gaa ggc agc gag cca gat ggc ttc tgg gag gcc ctg ggc ggg	2394
Val Ala Glu Gly Ser Glu Pro Asp Gly Phe Trp Glu Ala Leu Gly Gly	
635 640 645	
aag gct gcc tac cgc aca tcc cca cgg ctg aag gac aag aag atg gat	2442
Lys Ala Ala Tyr Arg Thr Ser Pro Arg Leu Lys Asp Lys Lys Met Asp	
650 655 660	
gcc cat cct cct cgc ctc ttt gcc tgc tcc aac aag att gga cgt ttt	2490
Ala His Pro Pro Arg Leu Phe Ala Cys Ser Asn Lys Ile Gly Arg Phe	
665 670 675	
gtg atc gaa gag gtt cct ggt gag ctc atg cag gaa gac ctg gca acg	2538
Val Ile Glu Glu Val Pro Gly Glu Leu Met Gln Glu Asp Leu Ala Thr	
680 685 690 695	
gat gac gtc atg ctt ctg gac acc tgg gac cag gtc ttt gtc tgg gtt	2586
Asp Asp Val Met Leu Leu Asp Thr Trp Asp Gln Val Phe Val Trp Val	
700 705 710	
gga aag gat tct caa gaa gaa gaa aag aca gaa gcc ttg act tct gct	2634
Gly Lys Asp Ser Gln Glu Glu Glu Lys Thr Glu Ala Leu Thr Ser Ala	
715 720 725	

aag cgg tac atc gag acg gac cca gcc aat cgg gat cgg cgg acg ccc	2682
Lys Arg Tyr Ile Glu Thr Asp Pro Ala Asn Arg Asp Arg Arg Thr Pro	
730 735 740	
atc acc gtg gtg aag caa ggc ttt gag cct ccc tcc ttt gtg ggc tgg	2730
Ile Thr Val Val Lys Gln Gly Phe Glu Pro Pro Ser Phe Val Gly Trp	
745 750 755	
ttc ctt ggc tgg gat gat gat tac tgg tct gtg gac ccc ttg gac agg	2778
Phe Leu Gly Trp Asp Asp Asp Tyr Trp Ser Val Asp Pro Leu Asp Arg	
760 765 770 775	
gcc atg gct gag ctg gct gcc tga ggagggggcag ggccccacca tgtcaccggt	2832
Ala Met Ala Glu Leu Ala Ala *	
780	
cagtgccttt tggaactgtc cttccctcaa agaggcctta gagcgagcag agcagctctg	2892
ctatgagtgt gtgtgtgtgt gtgtgttgtt tctttttttt ttttttacag tatccaaaaa	2952
tagccctgca aaaattcaga gtccttgcaa aattgtctaa aatgtcagtg tttgggaaat	3012
taaatccaat aaaaacattt tgaagtgtga aaaaaaaaaa a	3053
<210> 32	
<211> 1064	
<212> DNA	
<213> Homo sapiens	
<220>	
<221> CDS	
<222> (76)..(843)	
<400> 32	
atttggccct cgaggccaag aattcggcac gagggcacgg gctttccaca gcgcggggga	60
acgggaggct gcagg atg gtc aag ctg acg gcg gag ctg atc gag cag gcg	111
Met Val Lys Leu Thr Ala Glu Leu Ile Glu Gln Ala	
1 5 10	
gcg cag tac acc aac gcg gtg cgc gac cgg gag ctg gac ctc cgg ggg	159
Ala Gln Tyr Thr Asn Ala Val Arg Asp Arg Glu Leu Asp Leu Arg Gly	
15 20 25	
tat aaa att ccc gtc att gaa aat cta ggt gct acg tta gac cag ttt	207
Tyr Lys Ile Pro Val Ile Glu Asn Leu Gly Ala Thr Leu Asp Gln Phe	
30 35 40	
gat gct att gat ttt tct gac aat gag atc agg aaa ctg gat ggt ttt	255
Asp Ala Ile Asp Phe Ser Asp Asn Glu Ile Arg Lys Leu Asp Gly Phe	
45 50 55 60	
cct ttg ttg aga aga ctg aaa aca ttg tta gtg aac aac aac aga ata	303
Pro Leu Leu Arg Arg Leu Lys Thr Leu Leu Val Asn Asn Asn Arg Ile	

				65					70					75					
tgc cgt ata ggt gag gga ctt gat cag gct ctg ccc tgt ctg aca gaa	Cys Arg Ile Gly Glu Gly Leu Asp Gln Ala Leu Pro Cys Leu Thr Glu	80	85	90	351														
ctc att ctc acc aat aat agt ctc gtg gaa ctg ggt gat ctg gac cct	Leu Ile Leu Thr Asn Asn Ser Leu Val Glu Leu Gly Asp Leu Asp Pro	95	100	105	399														
ctg gca tct ctc aaa tcg ctg act tac cta agt atc cta aga aat ccg	Leu Ala Ser Leu Lys Ser Leu Thr Tyr Leu Ser Ile Leu Arg Asn Pro	110	115	120	447														
gta acc aat aag aag cat tac aga ttg tat gtg att tat aaa gtt ccg	Val Thr Asn Lys Lys His Tyr Arg Leu Tyr Val Ile Tyr Lys Val Pro	125	130	135	495														
caa gtc aga gta ctg gat ttc cag aaa gtg aaa cta aaa gag cgt cag	Gln Val Arg Val Leu Asp Phe Gln Lys Val Lys Leu Lys Glu Arg Gln	145	150	155	543														
gaa gca gag aaa atg ttc aag ggc aaa cgg ggt gca cag ctt gca aag	Glu Ala Glu Lys Met Phe Lys Gly Lys Arg Gly Ala Gln Leu Ala Lys	160	165	170	591														
gat att gcc agg aga agc aaa act ttt aat cca ggt gct ggt ttg cca	Asp Ile Ala Arg Arg Ser Lys Thr Phe Asn Pro Gly Ala Gly Leu Pro	175	180	185	639														
act gac aaa aag aaa ggt ggg cca tct cca ggg gat gta gaa gca atc	Thr Asp Lys Lys Lys Gly Gly Pro Ser Pro Gly Asp Val Glu Ala Ile	190	195	200	687														
aag aat gcc ata gca aat gct tca act ctg gct gaa gtg gag agg ctg	Lys Asn Ala Ile Ala Asn Ala Ser Thr Leu Ala Glu Val Glu Arg Leu	205	210	215	735														
aag ggg ttg ctg cag tct ggt cag atc cct ggc aga gaa cgc aga tca	Lys Gly Leu Leu Gln Ser Gly Gln Ile Pro Gly Arg Glu Arg Arg Ser	225	230	235	783														
ggg ccc act gat gat ggt gaa gaa gag atg gaa gaa gac aca gtc aca	Gly Pro Thr Asp Asp Gly Glu Glu Glu Met Glu Glu Asp Thr Val Thr	240	245	250	831														
aac ggg tcc tga gca gtgaggcaga tgtataataa taggccctct tggaacaagt	Asn Gly Ser *	255			886														
cttgcttttc gaacatggta taatagcctt gtttgtgtta gcaaagtgga atctatcagc					946														
attgttgaaa tgcttaagac tgctgctgat aattttgttaa tataagtttt gaaatctaaa					1006														
tqtcaatttt ctacaaatta taaaaataaaa ctccactcac tatqctaataa aaaaaaaaa					1064														



gaa gcc cgc tgt ttc tat ggc caa att gcc atg gaa aac ata cat Glu Ala Arg Cys Phe Tyr Gly Phe Gln Ile Ala Met Glu Asn Ile His 160 165 170	710
tct gaa atg tat agt ctt ctt att gac act tac ata aaa gat ccc aaa Ser Glu Met Tyr Ser Leu Leu Ile Asp Thr Tyr Ile Lys Asp Pro Lys 175 180 185	758
gaa agg gaa ttt ctc ttc aat gcc att gaa acg atg cct tgt gtc aag Glu Arg Glu Phe Leu Phe Asn Ala Ile Glu Thr Met Pro Cys Val Lys 190 195 200	806
aag aag gca gac tgg gcc ttg cgc tgg att ggg gac aaa gag gct acc Lys Lys Ala Asp Trp Ala Leu Arg Trp Ile Gly Asp Lys Glu Ala Thr 205 210 215 220	854
tat ggt gaa cgt gtt gta gcc ttt gct gca gtg gaa ggc att ttc ttt Tyr Gly Glu Arg Val Val Ala Phe Ala Val Glu Gly Ile Phe Phe 225 230 235	902
tcc ggt tct ttt gcg tcg ata ttc tgg ctc aag aaa cga gga ctg atg Ser Gly Ser Phe Ala Ser Ile Phe Trp Leu Lys Lys Arg Gly Leu Met 240 245 250	950
cct ggc ctc aca ttt tct aat gaa ctt att agc aga gat gag ggg ttt Pro Gly Leu Thr Phe Ser Asn Glu Leu Ile Ser Arg Asp Glu Gly Phe 255 260 265	998
aca ctg tga ttttgct tgcctgatgt tcaaacacct ggtacacaaa ccatcggagg Thr Leu * 270	1054
agagagtaag agaaataatt atcaatgctg ttcggataga acaggagttc ctcaactgagg	1114
ccttgccctgt gaagctcatt gggatgaatt gcactctaata gaagcaatac attgagtttg	1174
tggcagacag acttatgctg gaactggggt ttagcaagggt tttcagagta gagaacccat	1234
ttgactttat ggagaatatt tcactgggaag gaaagactaa cttcttttgag aagagagtag	1294
gcgagtatca gaggatggga gtgatgtcaa gtccaacaga gaattctttt accttggatg	1354
ctgacttcta aatgaactga agatgtgccc ttacttggct gatttttttt ttccatctca	1414
taagaaaaat cagctgaagt gttaccaact agccacacca tgaattgtcc gtaatgttca	1474
ttaacagcat ctttaaaact gtgtagctac ctcaacaacca gtccctgtctg tttatagtgc	1534
tggtagtatc accttttgcc aaaaggcctg gctggctgtg acttaccata gcagtgacaa	1594
tggcagtctt ggctttaaaag tgaggggtga cccttttagtg agcttagcac agcgggatta	1654
aacagtcctt taaccagcac agccagttaa aagatgcagc ctcaactgctt caacgcagat	1714
tttaatgttt acttaaatat aaacctggca ctttacaaac aaataaacat tgtttgtact	1774

cacaaggcga taatagcttg atttatattgg tttctacacc aaatacattc tcctgaccac 1834  
 taatgggagc caattcaciaa ttcactaagt gactaaagta agttaaactt gtgtagacta 1894  
 agcatgtaat ttttaagttt tatttttaatg aattaaaata tttgttaacc aacttttaaag 1954  
 tcagtcctgt gtatacctag atattagtca gttggtgcc a gatagaagac aggttgtgtt 2014  
 tttatcctgt ggcttgtgta gtgtcctggg attctctgcc ccctctgagt agagtgttgt 2074  
 gggataaagg aatctctcag ggcaaggagc ttcttaagtt aaatcactag aaatttaggg 2134  
 gtgatctggg ctttcatatg tgtgagaagc cgtttcatth tatttctcac tgtattttcc 2194  
 tcaacgtctg gttgatgaga aaaaattctt gaagagtttt catatgtggg agctaaggta 2254  
 gtattgtaaa atttcaagtc atccttaaac aaaatgatcc acctaagatc ttgcccctgt 2314  
 taagtgggta aatcaactag aggtggttcc tacaagttgt tcattctagt tttgtttgg 2374  
 gtaagtaggt tgtgtgagtt aattcattta tatttactat gtctgttaaa tcagaaattt 2434  
 tttattatct atgttcttct agattttacc tgtagtcat acttcagtca cccagtgtct 2494  
 tattctggca ttgtctaaat ctgagcattg tctaggggga tcttaaaactt tagtaggaaa 2554  
 ccatgagctg ttaatacagt ttccattcaa atattaattt cagaatgaaa cataattttt 2614  
 tttttttttt ttgaaatgga gtctcct 2641

<210> 34  
 <211> 554  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> (127)..(318)

<400> 34  
 cacgaattga atttaggtga cactatagaa gagctatgac gtcgcatgca cgcgtacgta 60  
 agcttggtatc ctctagagcg gccctactac tactaaattc gcggccgcgt cgacctgtga 120  
 agaaac atg gcg gcc gcg acg ttg act tcg aaa ttg tac tcc ctg ctg 168  
 Met Ala Ala Ala Thr Leu Thr Ser Lys Leu Tyr Ser Leu Leu  
 1 5 10  
 ttc cgc agg acc tcc acc ttc gcc ctc acc atc atc gtg ggc gtc atg 216  
 Phe Arg Arg Thr Ser Thr Phe Ala Leu Thr Ile Ile Val Gly Val Met  
 15 20 25 30  
 ttc ttc gag cgc gcc ttc gat caa ggc gcg gac gct atc tac gac cac 264  
 Phe Phe Glu Arg Ala Phe Asp Gln Gly Ala Asp Ala Ile Tyr Asp His



	35	40	45	
atc aac gag ggg aag ctg tgg aaa cac atc aag cac aag tat gag aac				312
Ile Asn Glu Gly Lys Leu Trp Lys His Ile Lys His Lys Tyr Glu Asn				
	50	55	60	
aag tag ttccttgag gccccatcc aggccagaag gaccaggtcc acccagcagc				368
Lys *				
tgtttgccca gagctggagc ctcagcttga agatgatgct caaggtactc ttcattggacc				428
accattcgct gttggcaaga aacggcttta cttacaaaac agactcttta cttctctgctg				488
tgtttgaagt atgttttagtc agcatgctca ggaaataaat gtgaattgcc cttgaaaaaa				548
aaaaaa				554
<210> 35				
<211> 2769				
<212> DNA				
<213> Homo sapiens				
<220>				
<221> CDS				
<222> (136)..(1035)				
<400> 35				
cagtggcgga attctagcca ccggtgtctc cttctacatc cgctctgcg ccggctgcca				60
cccgcgctcc ctccgccgcc gccgccttgc tgetgctcaa agctgctgcc gcccttggg				120
ctaaaagggtt ttcaa atg gaa cat ttt gat gca tca ctt agt acc tat ttc				171
Met Glu His Phe Asp Ala Ser Leu Ser Thr Tyr Phe				
1 5 10				
aag gca ttg cta ggc cct cga gat act aga gta aaa gga tgg ttt ctt				219
Lys Ala Leu Leu Gly Pro Arg Asp Thr Arg Val Lys Gly Trp Phe Leu				
15 20 25				
ctg gac aat tat ata ccc aca ttt atc tgc tct gtc ata tat tta cta				267
Leu Asp Asn Tyr Ile Pro Thr Phe Ile Cys Ser Val Ile Tyr Leu Leu				
30 35 40				
att gta tgg ctg gga cca aaa tac atg agg aat aaa cag cca ttc tct				315
Ile Val Trp Leu Gly Pro Lys Tyr Met Arg Asn Lys Gln Pro Phe Ser				
45 50 55 60				
tgc cgg ggg att tta gtg gtg tat aac ctt gga ctc aca ctg ctg tct				363
Cys Arg Gly Ile Leu Val Val Tyr Asn Leu Gly Leu Thr Leu Leu Ser				
65 70 75				
ctg tat atg ttc tgt gag tta gta aca gga gta tgg gaa ggc aaa tac				411
Leu Tyr Met Phe Cys Glu Leu Val Thr Gly Val Trp Glu Gly Lys Tyr				

				80				85				90								
aac	ttc	ttc	tgt	cag	ggc	aca	cgc	acc	gca	gga	gaa	tca	gat	atg	aag					459
Asn	Phe	Phe	Cys	Gln	Gly	Thr	Arg	Thr	Ala	Gly	Glu	Ser	Asp	Met	Lys					
95				100				105												
att	atc	cgt	gtc	ctc	tgg	tgg	tac	tac	ttc	tcc	aaa	ctc	ata	gaa	ttt					507
Ile	Ile	Arg	Val	Leu	Trp	Trp	Tyr	Tyr	Phe	Ser	Lys	Leu	Ile	Glu	Phe					
110				115				120												
atg	gac	act	ttc	ttc	ttc	atc	ctg	cgc	aag	aac	aac	cac	cag	atc	acg					555
Met	Asp	Thr	Phe	Phe	Phe	Ile	Leu	Arg	Lys	Asn	Asn	His	Gln	Ile	Thr					
125				130				135								140				
gtc	ctg	cac	gtc	tac	cac	cat	gcc	tcg	atg	ctg	aac	atc	tgg	tgg	ttt					603
Val	Leu	His	Val	Tyr	His	His	Ala	Ser	Met	Leu	Asn	Ile	Trp	Trp	Phe					
				145				150								155				
gtg	atg	aac	tgg	gtc	ccc	tgc	ggc	cac	tct	tat	ttt	ggg	gcc	aca	ctt					651
Val	Met	Asn	Trp	Val	Pro	Cys	Gly	His	Ser	Tyr	Phe	Gly	Ala	Thr	Leu					
160				165				170												
aat	agc	ttc	atc	cac	gtc	ctc	atg	tac	tct	tac	tat	ggg	ttg	tcg	tca					699
Asn	Ser	Phe	Ile	His	Val	Leu	Met	Tyr	Ser	Tyr	Tyr	Gly	Leu	Ser	Ser					
175				180				185												
gtc	cct	tcc	atg	cgt	cca	tac	ctc	tgg	tgg	aag	aag	tac	atc	act	cag					747
Val	Pro	Ser	Met	Arg	Pro	Tyr	Leu	Trp	Trp	Lys	Lys	Tyr	Ile	Thr	Gln					
190				195				200												
ggg	cag	ctg	ctt	cag	ttt	gtg	ctg	aca	atc	atc	cag	acc	agc	tgc	ggg					795
Gly	Gln	Leu	Leu	Gln	Phe	Val	Leu	Thr	Ile	Ile	Gln	Thr	Ser	Cys	Gly					
205				210				215								220				
gtc	atc	tgg	ccg	tgc	aca	ttc	cct	ctt	ggg	tgg	ttg	tat	ttc	cag	att					843
Val	Ile	Trp	Pro	Cys	Thr	Phe	Pro	Leu	Gly	Trp	Leu	Tyr	Phe	Gln	Ile					
				225				230								235				
gga	tac	atg	att	tcc	ctg	att	gct	ctc	ttc	aca	aac	ttc	tac	att	cag					891
Gly	Tyr	Met	Ile	Ser	Leu	Ile	Ala	Leu	Phe	Thr	Asn	Phe	Tyr	Ile	Gln					
240				245				250												
acc	tac	aac	aag	aaa	ggg	gcc	tcc	cga	agg	aaa	gac	cac	ctg	aag	gac					939
Thr	Tyr	Asn	Lys	Lys	Gly	Ala	Ser	Arg	Arg	Lys	Asp	His	Leu	Lys	Asp					
255				260				265												
cac	cag	aat	ggg	tcc	atg	gct	gct	gtg	aat	gga	cac	acc	aac	agc	ttt					987
His	Gln	Asn	Gly	Ser	Met	Ala	Ala	Val	Asn	Gly	His	Thr	Asn	Ser	Phe					
270				275				280												
tca	ccc	ctg	gaa	aac	aat	gtg	aag	cca	agg	aag	ctg	cgg	aag	gat	tga					1035
Ser	Pro	Leu	Glu	Asn	Asn	Val	Lys	Pro	Arg	Lys	Leu	Arg	Lys	Asp	*					
285				290				295												

aggacttctt	tcatcaattc	aaaaccccta	gaaaacgtat	acagattata	taagtaggga	1215
taagatttct	aacatttctg	ggctctctga	cccctgcgct	agactgtgga	aagggagtat	1275
tattatagta	tacaacactg	ctgttgcctt	attagttata	acatgatagg	tgctgaattg	1335
tgattcacia	tttaaaaaca	ctgtaatcca	aacttttttt	tttaactgta	gatcatgcat	1395
gtgattgtaa	atgtaaattt	gtacaatggt	gttatggtag	agaaacacac	atgccttaaa	1455
atttaaaaag	cagggcccaa	agcttattag	tttaaatag	ggtatgtttc	aagtttgtat	1515
taatttgtaa	tagctctgtt	tagaaaaaat	caaagaccat	gatttatgaa	actaatgtga	1575
cataatttcc	agtgacttgt	tgatgtgaaa	tcagacacgg	caccttcagt	tttgtactat	1635
tggttttgaa	tcaagcaggc	tcaaactctag	tggaacagtc	agtttaactt	tttaacagat	1695
cttatttttt	tatttttgagt	gccactatta	atgtaaaaag	ggggggggctc	tacagcagtc	1755
gtgatgaaac	ttaaatatat	attctttgtc	ctcgagattt	taggaagggt	gtagggtgag	1815
taggccattt	ttaatttctg	aagtgctaag	tgtttttata	cagcaaacia	aaagtcaatt	1875
ttgctttcca	ccagtgcgag	agaggatgta	tacttttcaa	gagagatgat	tgctatttta	1935
ccgtttgaca	gagtcccgta	gatgagcaat	ggggaactgg	ttgccagggt	ctaaatttgg	1995
attgatattat	gcactgttat	ctgttttgac	acagatttcc	ttgtaaaatg	tgcttagttt	2055
acaaaaatta	acaaaggggg	ggaaaggacc	ttagaacttt	ttaaggtaaa	atcaaatata	2115
gctacagcat	aagagaatcg	agaaatttga	tagaggtaac	ttgtttaatg	taaatctaata	2175
agtacttgta	atttctttct	gcttagaatc	taaagatgtg	tttagaacct	cttgtttaaa	2235
aataatagac	tgcttatcat	aaaatcacat	ctcacacatt	tgaggcagtg	gtcaaacagg	2295
taaagcctat	gatgtgtgtc	attttaaaagt	gtcgggaattt	agcctctgaa	taccttctcc	2355
attgggggaa	agatattctt	ggaaccactc	atgacatatc	ttagaaggtc	attgacaatg	2415
tataaactaa	ttgttggttt	gatatttatg	taaatatcag	tttaccatgc	tttaattttg	2475
cacattcgta	ctatagggag	cctattgggt	ctctatttagt	cttgtgggtt	ttctgtttga	2535
aaaggagtca	tggtcatctgt	ttacattttac	cttatcaaac	ctagaatgtg	tatatattata	2595
aatgtatgtc	ttcattgcta	ggtactaatt	tgcatgtgtc	tttacaatatt	tcaatacaga	2655
aactataaca	ttcaatagtg	tgctgtcaaa	gtgtgcttag	ctcacctgga	tataacctaca	2715
ttgttaaagt	tctaaacagt	aatcattaaa	acatttttga	ttaaaaaaaa	aaaa	2769